Charles University
Faculty of Science

Department of Philosophy and History of Science
Theoretical and Evolutionary Biology

Mgr. Silvia Boschetti

HUMAN PERCEPTION IN CONDITION OF UNCERTAINTY: THE VISUAL, AUDITORY AND EMBODIED RESPONSES TO AMBIGUOUS STIMULI

LIDSKÉ VNÍMÁNÍ V SITUACI NEJISTOTY: VIZUÁLNÍ, AUDITIVNÍ A VTĚLENÉ REAKCE NA NEJEDNOZNAČNÉ STIMULY

Doctoral thesis

Supervisor: Prof. RNDr. Jaroslav Flegr, CSc.
Consultant: Prof. Hermann Prossinger, Ph.D.

Prague, 2023

Prohlášení

Prohlašuji, že jsem předkládanou práci zpracovala samostatně a použila uvedené prameny a literaturu. Dále prohlašuji, že práce nebyla využita k získání jiného či stejného titulu.Současně dávám svolení k tomu, aby tato práce byla zpřístupněna v příslušné knihovně UK aprostřednictvím elektronické databáze vysokoškolských kvalifikačních prací v repositáři Univerzity Karlovy a používána ke studijním účelům v souladu s autorským právem.

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged. The thesis has not been accepted for any degree and is not concurrently submitted for award of other degree. Also, I declare that I agree with availability of the thesis in the appropriate Charles University library and through an electronic database of university publications in the repository of Charles University and used for study purposes in accordance with copyright law.

V Praze dne 23. 2. 2023 Silvia Boschetti

Acknowledgements

I would like to thank my supervisor, Jaroslav Flegr, for the support, help and collaboration.
I thank my consultant, Hermann Prossinger, for the invaluable advice and help he gave me during my studies and for the knowledge he shared with me.
A special thanks to Jakub Binter, for his constant support, advice and friendship, without which I would not have been able to complete my research.
I want to thanks my colleagues, Robin Kopecký, Jelena Příplatová and Konrad Talmont-Kamiński for providing helpful suggestions and insights in my research project.
I would like to extend my sincere thanks to Tomas Hladký and Denisa Průšová, who encouraged me and assisted me in every way they could.
I thank Stefano Parmigiani, to inspired me to become a better scientist.
I wish to thank my partner, Radomir Elias, as well as my family for their continuous support though all these years and the encouragement, patience and understanding they provided me with, especially in times of difficulty.

I also thank the grant agencies GAČR and GAUK for financial support during the realization of the many studies and the ensuing furthering scientific development. Namely, the projects:
GAČR: Behavioral and Psycho-Physiological Response on Ambivalent Visual and Auditory Stimuli Presentation (19-12885Y)
GAČR: The bounds of religious cognition: Study of cognitive and cultural factors in Czech and Polish societies (21-31059S)
GAUK: Cognitive Bias, Agency Detection and Pattern Recognition: an Intercultural Quantitative and Experimental Study of Supernatural Belief (1404120)

Abstract

In order to orient ourselves in the environment our senses have evolved so as to acquire optimal information. The optimization, however, incurs mistakes. To avoid costly ones, the over-perception of patterns (in humans) augments the decision making. I tested the decision-making in two modalities, acoustic and visual. A set of stimuli (using computer-generated graphics, based on output from a very good pseudo random generator) was produced: masks with a random pattern with varying degree of transparency over geometrical figures were used, followed by similar task that involved black and white high-contrast patterns. In both cases, I was able to find, using a Bayesian statistical approach, that the ability to detect the correct pattern presence (or lack thereof) was related to respondents' thinking styles, specifically Rationality and Intuition. Furthermore, I used ambiguous facial expressions, and accompanying vocalizations, of high-intensity affects (pain, pleasure and fear) and low-intensity (neutral and smile/laughter). My findings evidenced that the high-intensity facial expressions and vocalizations were rated with a low probability of correct response. Differences in the consistency of the ratings were detected and also the range of probabilities of being due to chance (guessing). When arousal was manipulated in the respondents by using the harmless, but reliable, method Cold Pressor Task, the ratings of unimodal stimuli shifted towards higher accuracy for facial expressions of pleasure and laugh, and decreased in accuracy for vocalizations of laughter and neutral speech; all only for male stimuli. When the two modalities are presented simultaneously and in congruence, the probability of correct ratings did not increase for high-intensity displays but the due-to-chance calculations showed that none of the displays were rated due to chance except for the intersexual assessment of pleasure. In other words, when responding to the other-sex stimulus of pleasure, the valence is guessed by the respondent. In the incongruent conditions (for which no correct rating can exist), I found that most of the decisions are based on the auditory modality, the visual one being suppressed. The one exception was the facial expression of pleasure combined with neutral speech; it resulted in cross-talk — namely, a doubly incorrect rating.

Abstrakt

Naše smysly se vyvinuly tak, abychom z okolního prostředí získávat optimální množství informací. Tato optimalizace ovšem znamená, že je třeba počítat s chybami. Proto, abychom předešli těm s významným dopadem, vyvinula se u člověka tendence k nadhodnocování významu vzájemných souvislostí (i ve smyslu vnímání vzorů a posloupností). Ve své práci jsem testovala schopnost vyhodnocování vizuálních a akustických stimulů. Za použití počítačové grafiky byl vyvinut soubor testovacích stimulů, kde bylo rozložení prvků určeno sofistikovaným generátorem pseudo-náhodných čísel. Tyto výsledné masky s různou mírou průhlednosti byly užity k překrytí geometrických tvarů. Podobného postupu bylo užito k vytvoření černobílých stimulů s vysokým kontrastem. Za použití metod bayesovské statistiky jsem nalezla vzájemnou provázanost schopnosti určit přítomnost vzoru (a její absenci) a stylu myšlení, specificky racionálního a na intuici založeného. Dále jsem pak použila nejednoznačné výrazy tváře a vokalizace vysoce intenzivních afektivních stavů (bolest a slast) a stavů nízké intenzity (neutrální výraz/promluva, úsměv/smích). Výsledkem je zjištění, že vysoká intenzita projevu je spojena s nízkou schopností respondentů správně vyhodnotit valenci vizuálních i akustických stimulů. Díky použitému statistickému přístupu jsem mohla také vyhodnotit, zda je výsledek konzistentní a zda je důsledkem hádání. Všechny vizuální stimuly byly vyhodnoceny formou hádání, naopak akustické stimuly nejsou vyhodnoceny hádáním. Po užití manipulace zvyšující vzrušení hodnotitelů, šlo o bezpečnou, ale spolehlivou metodu (Cold Pressor Task), pravděpodobnost správného hodnocení stimulů se zvýšila v případě hodnocení výrazů obličejů slasti a úsměvu mužů. Naopak ke zhoršení pravděpodobnosti správného určení valence došlo v případě smíchu a neutrální promluvy – opět při hodnocení mužů jako stimulů. Pokud jsou prezentovány souhlasné stimuly kombinující vizuální a akustickou složku, nedojde ke zpřesnění určení valence v případě stimulů spojených s vysokou intenzitou. Žádné hodnocení není hádáním s výjimkou slasti a to vždy opačným pohlavím (muži hádají u žen a naopak). Při prezentaci nesouhlasných vizuálních a akustických stimulů, kdy nemůže být hodnocení správné, je vizuální modalita potlačena akustickou. Jedinou výjimkou je kombinace výrazu slasti v obličeji a neutrální promluvy. Tento stimul je hodnocen jako negativní (tedy nesprávně pro každou z modalit).

Table of Contents

Chapter 1: Overview

The study of perception and signal processing in biology has always benefited from an interdisciplinary approach. The sensory perception of the organism is clearly adapted to its environment in order to gain relevant information to survive, to reproduce while the cost/benefit ratio is constantly being taken into account. One limiting factor of the information retrieved from the environment is the metabolic cost during information decoding (by the sensory organs) but also during its cognitive elaboration in the brain. At the same time, the success of the actions of the organisms strongly depends on the accuracy of their perception. Nonetheless, in the process of perception flaws occur, and the percept is not always in agreement with the stimulus that elicits it. For this reason, both the condition of uncertainty and the existence of ambiguous stimuli are particularly interesting fields of study: situations that force the individual to make a decision that is based on the cost of possible mistakes (false positive and false negative errors), and heuristics (long- and short-term experience, innate mechanisms, learned aversions, etc.).

Error Management Theory describes the implications of costly decisions and the reasons for over-perception of patterns as a safety-mechanism, albeit in a runaway manner. How much information and how many cues are sufficient to assess the stimulus correctly, and what constitutes a strong enough signal are interesting questions to pose.

Numerous insights from fields like physics and mathematics, but also psychology and neuroscience have helped in the development of new theoretical and methodological frameworks for the study of perception. The recent advances in technology and computational power can and does bring this interdisciplinary endeavor even further. Specifically, by using audiovisual media, computer-generated graphics, Bayesian statistical procedures (not so novel in some fields), and Artificial Neural Networks, we could devise novel testing procedures using ambiguous stimuli where we could rely on objective measures of control before the research was begun.

By definition, ambiguous stimuli do not have a unique interpretation and so the cues present in these stimuli can lead to different percepts. Therefore, in the pre-testing phase, it is not only impossible but also counterproductive to reach any version of an agreement (one of which is the procedure used by researchers to ensure low interrater variability) about such stimuli among human raters whether these are experts or naïve.

To test the perception of visual patterns, we created a set of stimuli in which we could control the randomness or non-randomness of each of the stimuli. The use of a state-of-the-art pseudo-random number generator (which, at this level of practicality, is truly random) gave us the possibility to statistically evaluate stimuli regarding the presence or absence of the randomness without the need for it to be evident to human rater and, to boot, to the experimenter. For the first time, to our knowledge, we were able to create sets of stimuli to study false positive errors in visual pattern detection (pareidolia) and false negative errors (apoidolia).

One study involved stimuli sets in color, created with the use of a commercially available, patented automaton. This automaton was able to generate a set of stimuli with a gradient from random to non-random with varying difficulty for the non-random ones. This was achieved by manipulating the transparency of the random mask with a pattern underneath; the latter consisting of uniformly colored geometrical shapes. By utilizing Dirichlet distributions, we were able to identify differences between subjects who accurately rated the randomness or non-randomness of the images and those who exhibited pareidolia and apophenia. Additionally, we found that both of these perceptual errors are related to thinking style, specifically the rational (analytical and logical operation) and experiential (operating through automatic and associative learning) approaches to interpreting one's environment. We confirmed these results in a follow-up study where we used black and white stimuli (to ensure that the contrast between the elements in the images was maximal). In fact, cluster analysis shows that in each case of the tested thinking styles (subscales of Rationality and Intuition), there are two statistically distinct clusters relating these thinking styles to the proportion of correct identifications of the patterns and errors (either false positive or false negative). We note that the relation is a nonlinear one and was detected by the autoencoder.

In a further study dealing with visually ambiguous stimuli, we tested the assessment of human expressions of affective states with high or low intensity. Indeed, recent studies on facial and vocal affective expressions showed that intense affective states (characterized by high arousal) are ambiguous stimuli to human perceivers and are more difficult to correctly assess in comparison with low-intensity expressions, an effect known as the *paradox of intensity*. Furthermore, in the case of facial expressions pain and pleasure, even the use of the muscula action units (FACS system) to describe the grimace is not sufficient to distinguish between the two expressions. To assess the difference, we used an AI application with a feature extraction algorithm resulting in identifiable differences between the two, confirming that enough cues are present in the images of the faces to discriminate pain expressions from pleasure expressions for the machine learning environment but insufficient for the human perception, which, we conclude, could not catch such cues. Indeed, the results of subsequent studies, involving human raters, showed that the ratings were not only inaccurate and inconsistent; oftentimes the raters were simply guessing (evidenced by a novel statistical method that can calculate the probability of guessing). Interestingly, when testing the matching emotional vocalizations, we found that the vocalizations of high-intensity affective states were rated with a low accuracy but the results were not due to chance: this means that, in contrast to the visual stimuli, the raters were convinced (not guessing) about their rating even if they were incorrect — a truly remarkable discovery.

Furthermore, we tested whether the integration of this information (facial expressions and vocalizations presented simultaneously) will increase the accuracy of the assessment of the ambiguous stimulus when the two modalities carried information regarding the same affective states (congruent condition) and, alternatively, if we observe possible cross-talk when they are presented with different displays (incongruent condition). The redundancy of information in the congruent condition did not decrease the perception ambiguity of the highly intensive affective states: the raters did not succeed in distinguishing between pain and pleasure (both were rated almost equally often positive and negative). However, we observed a change in the probability of guessing by the raters: they were more confident in their decision in the congruent condition than when rating only facial expressions but less confident in comparison with the vocalization rating condition.

In the incongruent condition, surprisingly, we found evidence of sensory cross-talk in only one case: when the visual stimulus depicted a neutral expression was paired with a vocalization of pleasure. In this last case, neither of the modalities (visual or auditory) was suppressed, but the two interfered with each other, resulting in an assessment that matched neither of the modalities. Most often, however, we observed that the auditory modality suppressed the visual one. This may be due to the fact that the ratings of the vocalizations were not due to chance but the facial expressions were. Further research using a methodology that prevents guessing (for example, by implementing an experimental set-up that rigorously enforces The Lady Tasting Tea randomization) will be necessary to understand this mechanism better.

Finally, we tested whether the inner state of the rater (specifically: arousal induction) has an impact on the assessment of the stimuli, either facial expression or vocalization. To do so, we used a safe but reliable procedure that increases the cortisol levels of the raters (called Cold Pressure Task) in an experimental group. The procedure involves immersion of one lower limb (crus) in the ice-cold water for ~90 seconds, leading to an increase in their arousal while rating the stimuli. Surprisingly, we observed a more accurate assessment than in the control group for facial expressions of laugh and pleasure, and a lower accuracy of assessment for vocalizations of laughter and neutral was only on stimuli depicting a male expresser.

This thesis contains numerous innovations in stimuli creation with increased ecological validity (1) pretesting the stimuli without using a control group, (2) rating procedures, (3) reduction of the possibility of obtaining the result due to chance (and exactly calculating its probability), (4) use of questionnaire data without index computation, (5) statistical evaluation without of null hypothesis statistical testing fallacies, and (6) initating arousal in laboratory experiments.

Furthermore, two questionnaires were standardized into Czech, and a further four translated to two languages, Czech and Italian.

Chapter 2 History of Signal and Cue

2.1 Signal and cue in ethology

Charles Sanders Peirce was a man of many hats: a chemist, a mathematician, a logician, a philosopher, and a founder of philosophical pragmatism. He is considered the founder of abductive reasoning: a rigorously combined formulation of mathematical induction and deductive reasoning. He is one of the originators of the use of logical operators in electrical switching circuits that were later applied in digital computers — a change that made science different forever.

Modern semiotics is to a large degree inspired by Peirce´s work. Unfortunately, a majority of his findings was not published and he is known primarily for the inspiration he provided to other thinkers. The most well-known are William James, a psychologist, Konrad Lorenz and Nico Tinbergen, both biologists (ethologists) and together winners of the Nobel Prize in Physiology and Medicine.

James developed his theory of emotions, in which he suggested that the physiological trigger reaction is what determines the cascading feeling of emotions — thereby labeling emotions and affects.

Lorenz, after reading Pierce´s work, developed his theory of animal communication. Specifically, he was inspired by Peirce in differentiating between purposeful and non-purposeful communication. The difference between the two is the intentionality in the communicated message. An example of purposeful communication would be a rattlesnake's tail: an anatomic device that is used to communicate a warning message to the recipient. In contrast, the buzzing of a mosquito is a non-purposeful communication; it is perceived as a message by the to-be-attacked host but it is without a communicative value on the part of the mosquito. And it may even provide information, such as location, that the mosquito would arguably avoid communicating. The rattle is a specifically evolved set of scales that signals to the undesired (by the rattlesnake) creature the fact that the very next behavior of the snake would be an attack if the intruder were not to move out of harm's way. The mosquito buzz perceived by the host is a disadvantage because it prevents a stealthy approach. The buzz cannot be suppressed; the mosquito´s wings need to move so that flight and hovering is assured.

To the ethological community, these two modes of communication are known as signals and cues. According to the signaling theory developed by Scott-Philips, signals are defined as "Any act (or structure) that (i) affects the behavior of other organisms; (ii) evolved because of those effects; and (iii) which is effective because the effect (the response) has evolved to be affected by the act or structure." And cues are defined as "Any act (or a structure) that (i) affects the behavior of other organisms; and (ii) which is effective because the effect has evolved to be affected by the act or structure; but which (iii) did not evolve because of those effects." (Scott-Philips, 2008, p.387).

Signals in a biological/ethological sense are very different from what is defined as a signal in the physical sciences. There are only four possibilities for the ethological concept of signal:

1) Nothing is sent — so nothing has been perceived.
2) A cue has been leaked from the sender and the receiver can perceive it and possibly adjust his/her behavior accordingly.
3) The sender has intentionally sent the signal and the receiver has perceived it.
4) The sender has sent the signal or cue but the receiver did not register it.

There is hardly any literature that would focus on when too much information has been lost for a signal to become a cue and when many cues are so reliably interpreted that they take over the role of being a signal.

2.2 Signal transmission, noise, entropy and all that

As opposed to the aforementioned signaling theory, the relation between communication, information and signal transmission in physics and the concept of entropy was proposed in 1948 by Claude Shannon in the paper *A Mathematical Theory of Communication*. Shannon was a mathematician working at Bell Laboratories in New Jersey, USA and was developing a theory to understand the loss of information during the transmission of messages — for instance electromagnetic signals traveling via some medium while two individuals were mutually sending and receiving messages, as during a telephone call.

Although his model was originally developed to deal with the observation that there is a loss of information during transmission, he realized that the problem of information loss applies to any type of information transmission. After transmission, the received signal it is not the same as what the sender sent. During the transmission the information decreased; the difference between the original information and the received information is called *noise*. I note that entropy is a quantity — and, therefore, so is information — as Shannon discovered in 1948. Indeed, both entropy and information can be measured; they both have the (physical) unit $J \cdot K^{-1}$. Just as it is not valid to refer to amount of entropy (or, for that matter, amount of energy), it is also invalid to refer to amount of information (and, consequently, amount of noise is also invalid). Information is additive, so information has been transmitted from two channels leads to more information in the sense that the information from these two channels can be added: $I_{\text{received}} < (I_{\text{source}_1} + I_{\text{source}_2})$. When the signal is corrupted during transmission, information is decreased (hence the "<" in the formula) and noise is increased.

The sum remains constant during this degradation process.

A naïve explanation of this phenomenon is that noise was added to the signal, changing the information and being a distractor when the receiver decodes the message. Shannon was well aware of the fact that the noise is not an *addition* that masks the information, but it is an actual *loss* of information.

He developed an equation that allowed him to calculate the loss. After sharing his results and insights with another mathematician, John von Neumann (the inventor of the hardware implementation of the Turing machine), suggested that this loss should be named *entropy*, because of the formula's identity with Boltzmann's formula for entropy in the statistical modeling of the 2nd Law of Thermodynamics. Later, both Shannon and van Neumann, and, ultimately, the whole physical science community observed that not only the mathematical expression for Shannon's and Boltzmann's entropy were the same, but the underlying mechanism was identical.

Boltzmann was a physicist working on the statistical interpretation of the kinetic theory of gases and he developed the formula for entropy, namely $S = k \ln\Omega$ (k is the Boltzmann constant and Ω is the number of microstates of a system). Boltzmann's discovery was that only if a system is understood statistically can experimental outcome fluctuations be understood and these can only be calculated with his formula. These fluctuations are the noise; noise is, therefore, the increase in entropy because of the 2nd Law of Thermodynamics makes the increase inevitable.

The distinction between information and noise is very important in the scientific study of perception. During the transmission of the message or signal, the information contained is deteriorated by the increase in entropy. The information, therefore, needs to be retrieved via decoding and reconstruction; the challenge is to retrieve enough of the original meaning.

Not only physicists and engineers were aware of the mechanism of information loss — even prior to Shannon's discovery of the relation of the loss with the 2nd Law of Thermodynamics. The problem the communication experts were trying to solve was how to ensure a minimal loss (preferably zero) due to the transmission. Their approach was to implement a redundancy so as to aid in reconstruction. A general, still widespread, misconception is to consider redundancy as a repeat of the signal a second time. This must be a false approach, because the receiver is not capable of, in the case of the two signals being different, knowing which of these two signals is the one with the smallest corruption. Repeatedly sending a signal does not eliminate signal corruption and signal loss.

The approach is to ensure redundancy by clever encoding schemes. Internal relations between elements of the signal can be encoded with an invertible encoding algorithm. The receiver inverts the encoding to retrieve the maximum possible information in the message.

Clearly, the amount of information transmitted via the signal has increased, because this redundancy has also been transmitted. However, in all these encoding/decoding schemes, the algorithm used for generating the redundancy is transmitted beforehand (again, obviously, via an encoded signal). The more secure the reconstruction of the sent signal, the more elaborate the redundancy algorithm has to be.

A tempting thought in this context is the speculation that the redundancy in biological signals is an algorithm that has evolved: maybe the decoding of the received signal is possible because 'evolution' has made it possible that redundancy is invertible.

This plethora of signal transmission, noise and redundancy features is often summarized as communication of information between two individuals: the sender and the receiver. Ethologists are not adequately rigorous in these aspects of signal transmission. Rather, they focus on the quality or reliability of the signal. In this thesis, I communicate to you, the reader, the interplay between perception, signal properties and noise by describing several of my published studies and possible conclusions and inferences for future studies. They show that unraveling noise and signal reliability is statistically very challenging.

2.3 Perception in biological descriptions

With the word *perception* we refer to the results of elaboration processes pertaining to sensory information. The word perception comes from the Latin *perceptio* (gathering, receiving, conceiving) and can be considered as the results of a process of organization and elaboration of the information that has been gathered through many sensory modalities.

It is essential to the survival of organisms for them to make inferences regarding the environment and their surroundings in order for them to perform any type of action. The success of the action will strongly depend on the accuracy (actually: reliability) of the perception. In the current scientific approach, two aspects of perception are of utmost interest: the sensory data (decoded by senses through sensory transduction) and their elaboration. Obviously, the accuracy of the perception is higher when the information is redundantly encoded (i.e. via congruent multimodal overlapping), allowing for a more robust percept. However, the metabolic cost of encoding stimuli limits this possibility. The encoding is, however, dependent on the species-specific *sensory dispositions*. These species-specific sensory dispositions refer to the results of an evolutionary sensory adaptation that allows each species to encode the type of environmental stimuli relevant for its survival.

The senses construct the species-specific perception of the environment. This concept of *Umwelt*, formulated by the Estonian biologist Uexkull, founder of biosemiotics, in 1920s is still relevant. The qualitative experience of the organism is fine-tuned via a species-specific, therefore subjective, construction and interpretation of the world that depend on the needs of the organism. Hence the species-typical *lived-world* (*Umwelt* in German). A quote by Uexkull

(1926, p.15) summarizes this situation:

> The task of biology consists in expanding in two directions the results of Kant's investigations: (1) by considering the part played by our body, and especially by our sense-organs and central nervous system, and (2) by studying the relations of other subjects (animals) to objects.

2.4 Perception in nonhuman animals

In octopi (*Abdopus aculeatus*), the lack of visual sensitivity to colors is compensated with the ability of detect polarized light (Temple et al., 2021). Light polarization is a property related to the plane of oscillation of the waves that compose light, in contrast to color which is a property related to the wavelength of these waves. In seawater the short and long wavelengths of visible light are more attenuated than the middle wavelengths, making color vision less reliable than polarized vision (Smith & Baker, 1981). As a result, polarized vision is more effective in detecting animals and inanimate objects in a marine environment and therefore allows for successful actions in the octopi's environment.

Differently from octopus, dogs (*Canis familiaris*) are adapted to a land environment. Their eye anatomy includes receptors that allow for color perception, namely a type of photoreceptor called cones that can transduce information regarding the wavelength emitted by objects. Their color vision is very different from ours. The spectrum of color that is visible to the subject is determined by the types of cones present in the retina of the eye. We, humans, have three different types of photoreceptors (for short wavelengths — violet and blue; for medium wavelengths — green, and orange; for long wavelengths — yellow and red), and the combined activity of these different types of cones allows us to perceive the full spectrum of the colors visible to the naked human eye. In the case of dogs, only two cones (for short wavelengths — blue and longer wavelengths — yellow) are present in the retina, and, consequently, their vision of the world is very different (Jacobs et al., 1993). The closest that a human can experience this type of color vision, is one type of color blindness, a genetic condition for which one or more types of cones do not function well — leading to lack of distinguishing between some wavelengths, and therefore not seeing some colors. To differentiate even further dogs' visual perception from ours: results of recent studies also suggest that dogs' vision is also sensitive to UV light (Byosiere et al., 2018).

Another animal with a color vision very different from ours is the mantis shrimp (*Stomatopod crustaceans*). It has 16 different types of photoreceptors; in addition to the ability of perceive UV light and circularly polarized light (Cronin et al., 2014), this animal needs only one eye to perceive spatial depth and it can move each eye independently. The structure of the mantis shrimp's eye is among the most complex in nature. Attempting to imagine how we could see the world with polarized light (i.e. the octopus) or with less colors (i.e. dogs) is difficult; it is certainly impossible to imagine how the world looks through a mantis shrimp's eyes.

2.5 Perception in Humans: Senses

Classically, the human perceptive senses are five: visual, auditory, olfactory, haptic (touch), and gustatory (taste). These senses have been recognized for millennia as the perceptual basis from which we create and infer the meaning of our environment (Aristoteles, 4th Cent. BC).

In reality, there are other sensory systems that may be less intuitive but also fundamental to our ability of perceiving and acting in the world. Depending on how we characterized a sense, up to 33 senses can be identified (Fairhurst, 2014). Considering the sensory modality as a type of transduction (meaning the transformation made by a receptor of various types of energy to electric signals that can be interpreted by the brain), we can identify the nociceptor (perception of pain), mechanoreceptor (proprioception, kinesthesis, balance, acceleration and muscle stretch), thermoreceptor (perception of temperature) and interoceptors (blood pressure and temperature).

We can extend this list further if we consider the specific sensation arising from the specific type of receptor. For example, in the case of vision, we can consider it as one sense (since it is based on the transformation of light energy to neurosignals), or as two (since we have two types of photoreceptors: the rods — sensitive to low light levels — and cones – sensible to the colors, but only at adequately high light levels) or as four (we have only one type of rods but three types of cones). In this way, humans can be considered to have from five to 33 senses.

Our conscious perception of the world takes into account that the information arriving via all these different channels creates at least one coherent output called a percept. Disentangling the role of the information carried by one channel becomes very complicated. Even more so when considering that even within the biological range of the senses of one single species, there is also an individuality aspect in the perception process, due to the elaboration of the information being related to the learning and the experiences of the individual organism.

2.6 Multimodal cues

The signals that a healthy individual receives from the different sensory channels need to be integrated in order to create one (coherent) information. How this information is integrated in terms of neural pathways and brain areas involved is a complex topic studied in numerous neuroscientific disciplines. In this present work, I focus on the behavioral consequences of multimodal perception, specific advantages in decision-making and the modality crosstalk issues.

Each individual needs to orient himself/herself in a highly complex environment and use information from the environment to regulate his/her decisions and behaviors. To this aim, the integration of information originating from different sources (in terms of sensory input) brings the advantage of a much more detailed assessment, in which many features can be taken into consideration (Metaxakis et al., 2018). Accessing different sensory inputs from the same stimulus can provide a better chance of correctly decoding the cues and the signal (Campanella & Belin, 2007). Indeed, the information carried by the different sensory modalities can be — in parts — redundant (in the sense of signal components from different sensory modalities carrying the same information or the same meaning; Akçay & Beecher, 2019). This can bring an advantage in terms of information decoding, decision-making and behavioral output, because the redundancy can compensate for the degradation of the signal (noise). In such cases, we can expect an improvement in terms of stimulus assessment and consequential behavioral choice. The opposite situation is when two (or more) senses bring two different information contents that are incompatible or incongruent with each other. We refer to such a phenomenon as modality crosstalk. As a result of such situations, the reliability of the assessment decreases. However, one of the incongruent sources of information could be suppressed, relying on the assumption that at least one other modality is considered to be more salient (Liu et al., 2019). The study of the effect of congruent or incongruent multimodal information can help to disentangle the role of the information carried by one sensory channel in the final percept and trigger of the consequential behavioral response.

2.7 Pattern perception

To detect meaningful cues and relationships between cues and other elements in the environment means to detect patterns. In this sense, humans are pattern-seeking animals that have developed cognitive and perceptual processes in response to evolutionary pressure — a pressure that facilitates the detection of patterns (Barrett, 2000). As theorized by the Error Management Theory (EMT; Haselton & Galperin, 2012), the cost-benefit relationship between false negative error (to not detect a pattern when it is present) and false positive error (to detect a pattern when it is not present) shows the possible advantage of overperception of patterns. To walk in a forest and miss-interpret a stick as being a snake (false positive error) has much lower cost than miss-interpreting a snake as being a stick (false negative error). The resulting bias in perceiving patterns when they are not present is called apophenia, and in the specific case of visual modality (illusonary visual pattern) is called pareidolia. The main feature of pareidolia is that random visual elements and noise are perceived by the individual as meaningful. Differently from visual hallucination (that arises without the need of a physical object), pareidolia is elicited by a real stimulus that is wrongly interpreted (Yokoi et al., 2014). The study of these phenomena can offer insight about which characteristics of the individual are associated with the ability to correctly distinguish between non-random patterns and random stimuli or, conversely, to become victim to a false positive or a false negative error. The subjective sensitivity to erroneous pattern detection may lead to another bias, the attribution of agency to non-living objects (Hyperactive Agency Detection Device; van der Tempel & Alcock, 2015), which leads, ultimately, towards supernatural beliefs (Barrett, 2000). Indeed, cognitive and perceptual biases can become candidates for hypotheses used to explain the emergence and the perseverance of religious and paranormal beliefs (Willard & Norenzayan, 2013). Previous studies have associated the overperception of patterns with different types of beliefs in the supernatural, as beliefs in the paranormal and believing the self-contradictory view that coincidences are not meaningless, but rather meaningful (Zhou & Meng, 2020; Bressan, 2002).

Most case studies of pareidolia are related to face perception. Elements in a stimulus that occupy the expected position of the eyes and of the mouth in the correct spatial arrangement are perceived as a face. However, faces are special stimuli for humans, because they are essential for our intra-specific communication. The relevance of the face as a stimulus has been documented in numerous studies (Tsao & Livingstone, 2008; Kanwisher & Yovel, 2006) and therefore some specific visual biases (i.e. face overperception) may be related to face pareidolia more than a generic pattern perception bias. For this reason, in our study we employed types of stimuli that are not derived from biological geometries and in which we can statistically control for the presence of patterns and noise.

2.8 Communication of affective states

The distinction between signals (in the traditional ethological sense) and cues is the results of ethological studies on animal communication (Lorenz, 1939) described above. The non-noise part of a signal is a physiological pattern that has evolved to actively convey specific information to a specific receiver. The success of the information transmission is related to the ability of the receiver to detect the signal, correctly interpret the meaningful part (decoding them reliably) and act consistently, based on the conveyed information. Signal transmission and signal detection are both evolutionary adaptations: both are fundamental parts of intra-specific and inter-specific communication.

Cues, on the other hand, transmit information but they did not evolve with the aim of communication. Both the cue(s) and the non-cue information contained in the signal can involve one or more sensory modalities, and both enable the organism to navigate in its environment.

In intra-specific communication, the sender encodes the information and transmits it to a receiver who decodes and interprets the message. For example, in human social interaction, the ability to read others' affective states is essential in order to, among other things, assess their mood and their mental state. To this goal, a variety of information is involved, ranging from non-verbal behavior (Grammer et al., 1998), facial (Grammer et al., 1990), and vocal expression (Leongómez et al., 2014). Darwin (1872) suggested, and later Paul Ekman tested (Ekman, 2006), that emotions and their facial and senso-motoric displays are universally recognized, as a sort of innate signaling system in humans. However, these claims were later criticized and several shortcomings were addressed in meta-analytical studies (Scherer et al., 2001; Elfenbein, et al., 2002). Generally, the decoding accuracy was highest (1) in naturally occurring emotional expressions (in line with the importance of ecological worth of the stimuli), (2) when raters and stimuli were of the same race, (3) when raters and stimuli had a similar cultural background (highlighting the role of previous experience of the decoders), (4) when participants were provided with visual and vocal cues simultaneously (Scherer et al., 2001; Elfenbein, et al., 2002) and (5) when the descriptors and the labelling of affective states are clear. For scientific research, there was the need to classify emotions as discrete or as continuous. The debate on this topic is still ongoing, with two main approaches: Discrete Category Theory (that considers each emotion as the result of a specific and distinct psychophysiological activation; Izard, 1994) and the Circumplex Model of Emotion (that considers emotion as a dimensional phenomenon; Posner at al., 2005; Russell, 1980). This second model suggests that all affective states arise from two mutually independent neurophysiological activations: represented geometrically by a finite valence interval (from extremely negative to extremely positive) and a finite arousal interval (from extremely low arousal to extremely high arousal). In this theory, affective experience is a linear combination of these two independent activations, which is then interpreted as representing a particular emotion (for review see Posner et al., 2005).

During the interpretation-decoding of the affective state that the sender is displaying, the context plays a very important role. By "context" we mean other information present in the environment that can be perceived and added to the original input. This increase can, on the one hand, be due to two stimuli detected with the same modality complementing their interpretative values (i.e. the two visual modalities facial expression and body posture) or, on the other hand, be detected with different modalities that can be employed to increase the information and, consequently, the accuracy of decoding. When a receiver can obtain information through visual (facial expression) and auditory (appropriate vocalization) channels, an approximately 90–100 % accuracy rate in the judgment of basic emotion perception (e.g., disgust, fear and happiness; Aviezer et al., 2008; Sourina & Liu, 2011) can be attained. In this case, the context, which is an enhancement due to additional cues available to the rater, adds information and increases the probability of correct decoding. However, the context may also consist of incongruent information that actually decreases the chance of correct assignment of the emotional state. If raters are presented with congruent information (i.e., emotional expression of the face and emotional prosody of the voice that match in terms of the emotion expressed), then the rate of correct identification is greater and response times are reduced (Wittfoth et al., 2010; Dolan et al., 2001). If the information provided is incongruent then confusion between signals occurs; ambiguity results. Since the two sources also cues regarding different states, we can suspect either (1) the suppression of one source in the interpretations or (2) the phenomenon of crosstalk, in which the two sources interfere with each other to the point that none is correctly identified. The discrepancy between incongruent stimuli is strongest when one modality represents a neutral state (Müller et al., 2011; Boschetti et al., 2023c).

Chapter 3: Laboratory and Stimuli Preparation Methodology

3.1 Stimuli construction in a laboratory

The use of a laboratory in experiments about human perception allows researchers to control many properties of the environment (i.e. illumination, sounds, and temperature, but also the display resolution of and the distance to the monitor) that would otherwise be confounding variables and possibly bias the data collected. Our use of the laboratory became important not only for the collection of the data but also for the construction of stimuli and their preparation. The stimuli were often first generated in a laboratory and then further refined in an attempt to reduce the corruption of the signal (van der Zant & Nelson, 2021). As explained above, the noise (entropy) is included in the transmission and the signal's degradation, therefore, can never be reduced during transmission. However, inclusion of redundancy in the transmission process can enable overcoming some of the degradation and thus increase the reliability of stimulus identification — at the cost of increasing the load to be transmitted along the communication channel.

In the study of affective states communication, databases created in laboratory are available and are used in many studies (Ma et al., 2015), increasing the possibilities of outcome comparisons. The images of these facial expressions are standardized in terms of background, face angle, eye level, illumination etc. (Lundqvist et al., 1998). Despite the attractiveness of making larger and larger data bases available, there are two shortcomings in this arrangement. First, a typical feature of these stimuli sets is that they were created by employing actors and actresses to produce the different requested facial expressions, namely those corresponding to mental images of different emotional states. Second, to confirm the communicative value of these stimuli, the images undergo a pre-test procedure before being used in a study. The pre-testing procedure involves a small number of test individuals who provide their ratings; these are used to compute inter-rater reliability. Those stimuli that best represent the desired expression are later used in the research.

On the one hand, this procedure provides results that are stable, replicable and allow for controlling confounding variables. On the other hand, this procedure increases the distance between the responses acquired in a laboratory experiment and the responses in an ecological valid situation, let alone in the real world. Stimuli produced by actors and actresses who emulate the emotions via facial expressions and related vocalizations have the advantage of being controllable and standardized. However, there is the risk that they are not perceived by the study participants in an adequately similar way as genuine expressions, because they are not elicited in the actors or actresses by the same physiological activation. For example, when a real, genuine laugh is compared to the acted laugh, the two are not only perceived differently in terms of categorization but they also activate distinct brain regions (McGettigan et al., 2015). There are instances where the specificity of the task demands the use of genuinely elicited affective expressivity. There are methods to produce the stimuli in more realistic ways: make the individuals undergo a procedure that causes them to experience the affective state (Sebe et al., 2007), or use pre-existing recordings of ecological valid situations and extract the parts containing the desired stimuli (Fernández-Dols at al., 2011; Aviezer et al., 2012; Wenzler et al., 2016; Raine et al., 2017; Boschetti et al., 2022; Binter et al., 2023a).

Another important issue deals with the emotional condition of the raters. In a real life scenario, the raters (or some other receiver of the stimulus) is not in a neutral emotional state when assessing the meaning of the stimulus. Instead, often he/she is in the same contextual situation that elicits the stimulus from the sender; therefore, the rater is not in a neutral condition. In addition, the sender may by affected emotionally by the stimulus, creating a circular communication of affective states that affect one another and modulate one another's interpretation. It is, of course, of interest to investigate how the manipulation of the affective states of the receiver may bring about insights in the assessment and interpretation of cues involved in different emotional states.

3.2 Inter-Rater Agreement

To confirm the communicative value of the stimuli, there are different pre-testing procedures available. One way of pre-testing the stimuli that depict emotional expressions is the inter-rater agreement (Ma et al., 2015). Before being used in an experimental set-up involving participants, the stimuli undergo a rating procedure with both naïve and expert raters who together categorized the emotion in the images or in the vocalizations' audio tracks. The resulting inter-rater agreement can be statistically derived as deviations (but not the differences, because the ratings are categorical variables) in the responses of the raters (Boschetti et al., 2022): Another possibility is to reach an agreement through an open debate about those stimuli that were rated or categorized differently (Robertson et al., 2010). This last approach is more often used when the raters are experts. There are several problems inherent in this procedure: (1) cultural and linguistic variations can have a strong impact; (2) the type of rating employed (categorization of the specific emotion expressed vs. categorization of valence or arousal of the expression); (3) the most relevant (in the context of this thesis) problem is that facial expressions of some type of emotions are inherently ambiguous (i.e. pain and pleasure; Wenzler et al., 2016). In these cases, arriving at an inter-rater agreement is not only very difficult but attempting to find an agreement can also be counterproductive, because there is the risk of choosing the images with a lower degree of ambiguity and, therefore, through the selection, constructing artificial prototypes that are even more distant from an ecologically valid situation.

3.3 FACS

Another way to pre-test and even create the facial expression stimuli is through the Facial Action Coding System. This system, originally proposed by Carl-Herman Hjortsjö (1970) and later developed by Ekman and Friesen (1978), is based on the fact that facial muscles are activated during emotional expression. There are various advantages of using this system, one of which is that instead of the grappling with the nature of the emotional category, it is the muscle activation that differentiates the emotions. In this way, the problem of language and culture in the labeling of emotions can be avoided (Hamm et al., 2011).The basic units of this system are the Action Units (AU), which correspond to specific muscle activations (i.e. upper lip raiser, brow lowerer, etc). The AU are used to catalogue the facial muscle group activations, expression by expression. Specific groups of AU activations correspond to specific emotional categories (Aviezer et al., 2015). Based on the expected AU activation for the specific emotion, it is possible to use FACS to pre-test the stimuli. Another advantage of this system is that the recognition of the AUs activated in the specific facial expression can be done by both trained human experts and specifically developed, dedicated software. Using the software, the information about the AUs activation can also be used in a prescriptive way, namely for the creation of animations of facial expressions. A third advantage of this system is that ambiguous emotional expressions can be straightforwardly described through the AU, without the need to reduce the expression to a prototype. With this (third) advantage, it was possible to demonstrate a further ambiguity involved in the facial expression of pain and pleasure: the AUs activated in these two cases are indeed extremely similar (Fernández-Dols et al., 2011). Using these methods could, therefore, not be sufficiently adequate to pre-test and distinguish between these types of stimuli.

3.4 AI as an evaluative tool

An alternative method to pre-test the visual stimuli is to implement Artificial Intelligence to quantify the differences between images. This approach allows the investigation of the differences in facial expressions without any prior categorization of muscle activation or rater input. Even when using software for muscle activation recognition and group identification of the AU involved in the expression, there is still a small number of differences (or cues) that the software considers relevant (Hamm et al., 2011). In contrast, the use of AI does not limit the analysis to the possible AUs involved. In the context of differentiation between facial expressions, applying a feature extraction function on the stimuli allows for the high-dimensional feature vectors that contain more information about the stimuli than AUs possibly can. These feature vectors are then dimension-reduced via an autoencoder (a neural network; Fig. 3-1), making it possible to calculate the Euclidean distance between pairs of these dimension-reduced feature vectors.

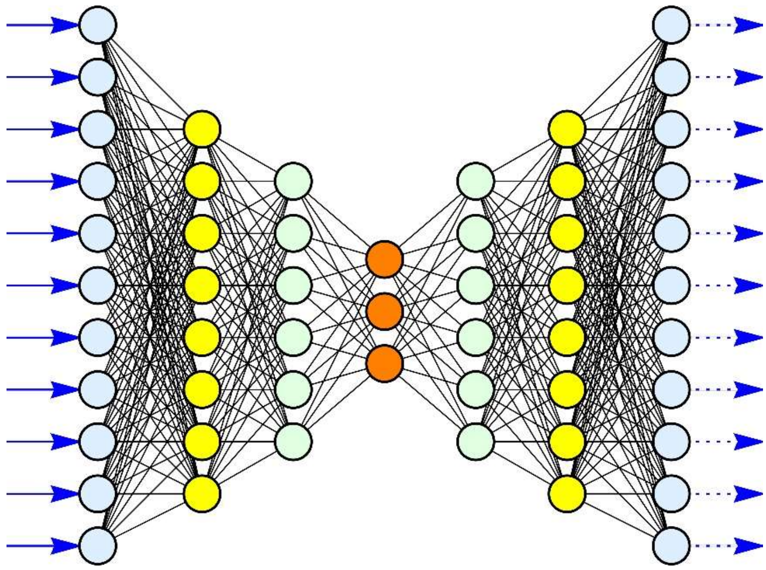


Figure 2-1 A symbolic rendition of an autoencoder. The 11 inputs (in the drawing) are represented by blue arrows from left to right. The inputs are conventionally labelled neurons (hence the name "neural network"). Each input neuron (light blue) has as many outputs as there are neurons in the next layer (8 in the drawing, represented as yellow discs). Each 'yellow' neuron thus has 11 inputs (represented as thin black lines, called edges). Each neuron in the yellow layer has as many outputs as there are neurons in the next layer: 6 outputs for each 'yellow' neuron and therefore 8 inputs for each 'green' neuron of the next layer. And so it continues: each 'green' neuron has as many outputs as there are neurons in the next layer (consisting of two 'orange' neurons) and each 'orange' neuron has 6 inputs. The light blue neurons on the left are called the input layer, the light blue neurons on the right are called the output layer. The (two) yellow layers, the (two) green layers and the (one) orange layer are called the hidden layers. An autoencoder always has the same number of output neurons as it has input neurons. The number of hidden layers, as well as the number of neurons in each hidden layer, is part of the design by the engineer constructing the autoencoder. The numerical values along the black edges between neurons are determined by an algorithm. The autoencoder attempts to produce an output equal to the input (hence the name 'autoencoder') without being an identity mapping. An important feature for modern autoencoders is the ability to cut (set to zero) certain interconnections (edges), or make them numerically very small (usually by using a sigmoid function). The central layer is called the code. If the inputs are the feature vectors, then the numerical values of the code are the components of the dimension-reduced feature vector. In the drawing: the input feature vector has 11 dimensions and the dimension-reduced feature vector has 2 components. Mathematically: if this is a successful autoencoder, it has detected nonlinear combinations between the components of the (input) feature vector that can be represented by two variables in a 2D space.

The Euclidean distance metric is extremely beneficial since it does not limit the dimensions of the feature vectors between which their distance is computed. This makes the tool versatile and multi-purpose.

One statistical analysis approach is to estimate of the ML distribution of the Euclidean distances and then calculate the modes and the $HDI_{95\%}$ confidence intervals (the Highest Density Interval at 95% significance; Kruschke, 2015). If one is interested in comparing the ML distributions of two such feature vector samples, it is possible to use a statistical test, such as the Wilks Lambda, to determinate whether they are significantly different. The use of this algorithmic approach was successfully employed to distinguish between the ambiguous facial expression of pain and pleasure (Prossinger et al., 2021b), proving that there are indeed objective differences between the two expressions. Further details and alternative approaches are discussed in Section 3 ("Statistical Methodologies").

One of our published studies (Prossinger et al., 2022a) precisely enumerated the far-from-trivial steps necessary for correct classification of pain and pleasure, which cannot be expected to be achieved by humans. Our results showed four clusters and two isolates. These clusters were detected after noise removal by the algorithm. The discovery of the necessity of noise removal provides further support for the two main arguments about the human inability to correctly rate the differences between pain and pleasure. First, the inter-individual facial expression variations are considerable, yet the noise corrupts the signal. Indeed, even though healthy individuals are equipped with facial muscles essential for basic emotion expression and the variability of the muscles involved is minimal (Waller et al., 2008), there are many influences related to the uniqueness of each individual´s expressions and limitations in their identification for other individuals in real world scenarios. Some of these limitations are: the fact that people choose or need to wear spectacles, some have beards, some are adorned with jewelry or with expensive makeup. All may these obstruct or alter the assessment of the facial expression. Further complications may arise in individuals who experienced facial nerve-related disorders or other central nervous system damage (Hamm et al., 2011). Furthermore, expressers' age-related features, their fat layer distribution, their skin texture, their general degree of facial expressiveness, and the morphology of their facial muscles are known to impact the production of their facial expressions (and consequently the probability of correct identification). We investigated identity uniqueness by using AI algorithms, because, *possibly*, individual expression familiarity potentially increases the accuracy of correct expression estimation by other humans if they are exposed to an individual for an adequately long, yet unknown in extent, period of time.

An interesting next step would be to test such a proposed explanation. Previous studies within this familiarity framework have been conducted on sadness, anger, and happiness; the results are mixed (Zhang & Parmley, 2015). In children, research on pain vocalizations has been published (Corvin et al., 2022); the study claimed that learning is the mechanism for obtaining proficiency with respect to specific expressers. It is worthwhile to compare how successful individuals are in assessing (rating) their partners and relatives in extremely (non-sexually) arousing moments (such as in sports encounters) with the ratings of strangers' facial expressions.

3.5 Use of ambiguous stimuli

Ambiguous stimuli are stimuli that do not have a clear interpretation and can trigger different cognitive elaborations and responses that depend on many influences, including individual differences, inner states of the receivers, and previous priming.

In ambiguous stimuli, the presented cues are conflicting among themselves or can be interpreted in different ways. This doesn't necessary imply that the stimulus does not have an objectively correct interpretation, but that the presence of the specific propriety in the stimulus or of specific cues make other interpretation as comparably probable as the correct one. When such stimuli are rated, the subjects need to resolve the perceptual ambiguity and then decide. This decision made not only based on the physical properties of the stimulus but also on the prior knowledge, experiences and expectations of the senders (Pollak et al., 2009). For this reason, ambiguous stimuli can be used not only to uncover mechanisms of decision making in conditions of sensorial uncertainty but more generally to help clarify the role of the physical proprieties of the stimulus and of the receiver's expectation. It is also important to mention that one of the possible decision-making mechanisms is guessing: if the ambiguity of the stimulus was not resolved the subject may guess; the response need not be based on the available information, past experiences, and risk — it can be a guess (Boschetti et al., 2022, Binter et al., 2023a).

3.6 Illusion

Illusions are specific types of ambiguous stimuli. In these cases, the interpretation of the information extrapolated from the stimulus is incorrect and results in an illusionary perception different from the physical reality (Walker et al., 2019).

We can distinguish between two main types of sensory illusion: physiological illusions are caused by physiological mechanisms while cognitive illusions are caused by some higher decoding process (Gregory, 1997). This distinction may appear to be artificial since these mechanisms affect each other and there are arguments for circularity and non-linearity of the process of perception being possible (Deneve & Jardri, 2016).

However, the distinction may still be useful to describe the role of sensory or cognitive functions in the case of a specific illusion. In the context of the research presented in this work, the physiological illusions are more related to physiological proprieties of the sensory systems and less to cognitive decoding attempts. An example is color after-effects, which consist of the illusionary colors perceived in a black and white image after a prolonged and static exposition to an inverse image of the one presented. This phenomenon is mainly related to the physiological proprieties of the photoreceptors, specifically the adaptation of the cones in the retina after the static exposition (Williams & MacLeod, 1979).

In comparison, cognitive illusions are related to the process of elaboration of the sensory information and to the interpretation of the presented stimulus. An example of this mechanism is so-called paraidolia, which is characterized as the illusionary perception of a pattern or a meaningful connection in (objectively) random stimuli (Merriam Webster Online, 2022). This type of illusionary perception is thought to be an adaptive property of perception, since the evolutionary advantage that this overperception is its proclivity to detect threats in complex sensory environment (Barrett, 2000). Indeed, it could be the result of an Error Management Theory (EMT) effect in the perception: detecting a pattern where there is none is less costly than non-detecting a pattern where is one (Johnson, 2009).

The specific case of face-pareidolia is the most studied phenomenon, and it is the illusory perception of a facial features (i.e. a face in the cloud or the face of Jesus on a 'burned' toast). There are certain features in the seen object that are distributed in a way that reminds the seeing observer of features of a face (i.e. two elements in the upper half — the eyes — and a larger element in the middle of the lower half — the mouth; a so-called three-point schema).

Since face is an evolutionarily relevant stimulus for humans, we are extremely primed to its perception (Kanwisher & Yovel, 2006; Tsao & Livingstone, 2008), to the point that we have an area of the brain dedicated primarily to face perception (Fusiform Face Area; Kanwisher & Yovel, 2006). Already a newborn orients its attention preferably to a face rather than to other stimuli and they can use information to distinguish between different faces a few days after birth (Field et al., 1984). In case of face-pareidolia, the brain elaborates the cues present in the images and guesses the most likely (yet incorrect) interpretation. However, the results obtained by studying face-pareidolia are not generalizable to other pattern perceptions.

Therefore, alternative approaches using different types of stimuli, such as computer-generated graphics depicting (environmentally relevant) patterns, are beneficial to uncovering more general mechanisms related to pattern perception. Furthermore, computer-generated graphics allows control over the presence or absence of patterns in the image with a certainty that can be mathematically evaluated.

3.7 Facial expressions and vocalizations of extreme intensity

As mentioned above, highly intensive affective states and their displays can be considered naturally occurring ambiguous stimuli. To date, these were rarely used for research purposes.

Among the naturally occurring ambiguous facial expression, it has been found that naturally occurring expressions of people in either a very highly positive or a very highly negative emotional situations are not easy to correctly rate along the positive-negative spectrum axis (Avierez 2012). For example, two completely opposite emotional affections — facial expressions of winners and losers in sports competitions (Aviezer et al., 2012), and more importantly, facial expressions of sexual gratification and pain (Hughes & Nicholson, 2008) were rated as indistinguishable. The phenomenon of affective ambiguity has only recently gained attention in the psychological literature, especially in the field of affective state processing. Recent studies (Aviezer et al., 2012; Blakeslee, 2006; Holz et al., 2021) argue that not only facial expressions but also vocalizations during sexual activities are indistinguishable from those of “suffering intense pain and agony” (Kinsey, 1953). A recent paper (Holz et al., 2021) refers to this phenomenon for which the vocalizations of intense affective states are more difficult to correctly attribute and categorize than low intensity affective as the “emotion intensity paradox.” This paradox exists, it is posited, due to a lack of further information that would allow a correct attribution and would probably be less likely to occur when all contextual information is available (Aviezer et al., 2012).

Chapter 4: Statistical Methodologies

4.1 The use of a rating scale

The pre-identification of categories (i.e. requesting the respondent to choose whether the emotion displayed is anger, fear or surprise) is more likely to capture the complex psychological representation of the emotion yet increase the variance — because it could then be uncontrollably affected by culture and linguistic differences — as criticized by Boschetti at al. (2022). The focus of previous studies was often on the emotions and affects as categories (without attention to how intense these emotions are) or on the dimensions of the emotions (high vs. low arousal or positive vs. negative), without categorizing the emotions. Consequently, outcomes of any study that avoids these limitations are very difficult to contrast with previous research that focused on the universality of specific categories (such as basic emotions) but not on others (the secondary emotions — the affects). Using stimuli labeled (during a pre-test) as pleasure or pain would inherently lead to testing whether participants agree on representations of pain and pleasure (that is to say, whether there is a common mental representation, as discussed by Chen et al., 2018).

4.2 Statistical Analysis of Responses

In field studies involving questionnaires, the responses are categorical variables, which may or may not be ordinal. The responses to the queries are ordinal numbers, not cardinal numbers (Blalock, 1960). They may not be directly converted into cardinal numbers, because a change in the choice of the mapping results in a change the statistical signal. (The statistical analysis is then the statistical analysis of the mapping, not of the data). Rather, the responses must be mapped into unit vectors. If, for a query with five response options, the chosen response is 'B', say, then the vector is $(0 \quad 1 \quad 0 \quad 0 \quad 0)^T$. If a respondent identifies a pattern as random, the response is $(1 \quad 0)^T$, and if it is non-random, then the response is $(0 \quad 1)^T$. The response vectors are then concatenated; the resulting vector called a feature vector (Murphy, 2012).
A detailed example: If there are two responses to (biological) sex (F/M), six responses to municipality (rural, small village, large village, small town, large town, city) and four to legal partnership status (single, married, divorced, widowed), then the encoding of biological sex is a vector in 2D, municipality is a vector in 6D, partnership status is a vector in 4D. The concatenated vector will have $2 + 6 + 4 = 12$ components, and only three of these will be nonzero (1, in fact) and these nonzero components will not be in arbitrary positions. A person may be female ($(1 \quad 0)^T$), live in a large village ($(0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0)^T$), and be divorced ($(0 \quad 0 \quad 1 \quad 0)^T$). Then the concatenated vector will be

$$(1 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0)^T.$$

Dimension-reduction of these feature vectors (one for each member in the survey) will be described in the next section. We note that the norm of the feature vector for n responses is $\sqrt{n}$.

4.3 Dimension Reduction and the Curse of Dimensionality

In any multivariate statistical analysis, one has a vector of realizations of many statistical populations, one per component. They are very rarely the same for each vector component. As an example, consider a field worker collecting the following data: (1) mass, (2) age, (3) height, (4) biological sex, (5) gender identity, (6) characterization of municipality, (7) sexual relationship, (8) citizenship, (9) voting choice, (10) emotional response to a political issue. Of these 10 vector components, (1)–(3) are cardinal numbers, noting each of their distributions are not normal (for mass, age, and height). The components (4)–(10) are categorical, which can be encoded with ordinal numbers (but need not be: see below); their distributions are Dirichlet distributions (or Beta distributions if the responses are binary). In this example, each datum is a vector in 10D. The obvious goal of a researcher (for this data set, presumably a polling analyst) is to characterize the data set with histograms (for cardinal numbers) and bar charts (for ordinal numbers); such characterizations of a data set are (*arguably*: despicably) called bean-counting.

Bean counting is useful (for the detection of data collection errors, for providing an overview, and many other reasons), but certainly not more than the beginning of a statistical analysis. Most obvious is the intent to determine possible relations among the components of the vectors — this can be, and most often is, very difficult.

One reason for the difficulty is the curse of dimensionality. Every vector component of the data set has a dispersion estimated by the variance. Any computation of variance is quadratic in the variable values, so the variance can only increase with an increase in sample size and in an increase in the number of variables being measured (or canvased, in the case of categorical variables). This phenomenon is aptly called the curse of dimensionality. Part of the curse is the difficult-to-interpret bean-counting graphs. The major part is, of course, the challenge of detecting the signal buried by the variance.

The most modern method of extracting the (statistical) signal from a multi-dimensional data set is the application of an autoencoder (Fig. 3-1), which is a special ANN (artificial neural network).

4.4 Clustering Algorithms

The Fig. 4-1 shows the output of the result of a dimension-reduction with an autoencoder. The dimension-reduced feature vectors (displayed as points in a plane, because the autoencoder succeeded in dimension reduction to 2D) are not uniformly distributed. The statistical/algorithmic challenge is to find clusters. The logic: if the points are not uniformly

distributed, can one then find points that are closer (in some sense) or more related to some other points, while at the same time not related to other points?
This question can be answered with clustering algorithms. These algorithms attempt to find clusters; the user specifies the type of relatedness (such as Euclidean distance function or connectivity or local point density). Most 'useful' (and preferred by statisticians) clustering algorithms do not specify a priori the number of clusters the algorithm is to find. Clusters are sometimes (especially in applications) called groups and the application of clustering algorithms is sometimes called partitioning the data set.

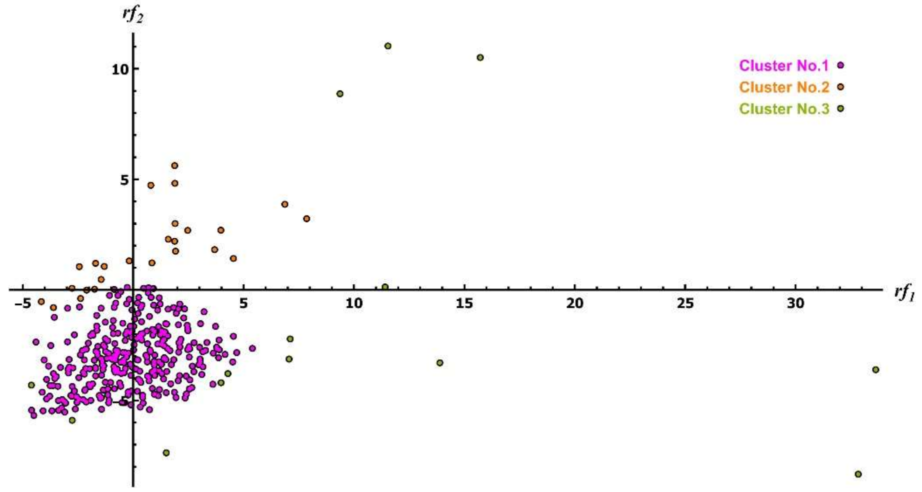


Figure 4-1 The distributions of clusters of the dimension-reduced feature vectors of the color boundaries chosen by the participants (Prossinger et al., 2023). The clusters were found using the DBSCAN algorithm.

The dimension reduction achieved by the autoencoder need not be to 2D vectors. Fig. 4-2 shows an example where the dimension-reduced vectors are in 3D.

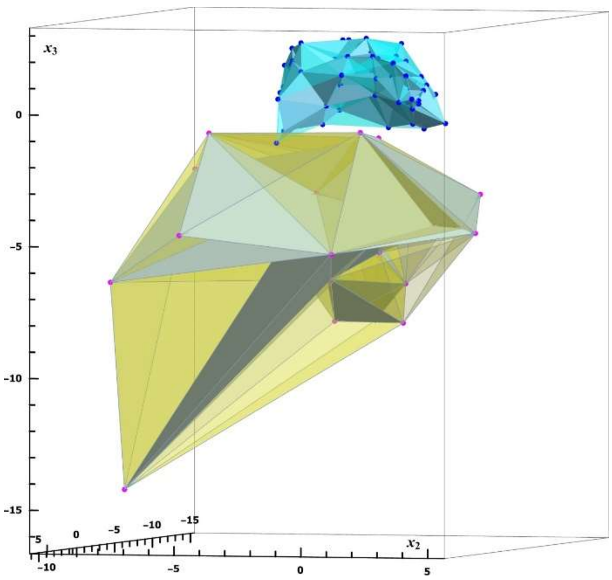


Figure 4-2 The two clusters obtained after dimension-reducing the feature vectors using an autoencoder. The points are the coordinates of the dimension-reduced feature vector of each participant's responses to

the Rationality queries together with the identification of a random or nonrandom distribution of squares. The concave hull connects the points in each of the clusters. For Rationality, Cluster #2 has 39 points and Cluster #1 has 174 points. (The results presented here are from the publication Boschetti et al., 2023b.)

Of the numerous clustering algorithms that have been developed in the recent past, here is a list of many that are often used: (1) DBSCAN, (2) spectral, (3) Jarvis-Patrick, (4) agglomeration, (5) spanning tree, (6) neighborhood contraction, (7) Jarvis-Patrick, (8) Gaussian mixture, (9) *K*-Means, (10) *K*-Medoids, and (11) mean shift. The algorithms (9) and (10) require a pre-specification of how many clusters are to be detected. They are to be used only in very special circumstances, such as when preprocessing data for neural networks used as classifiers or for SVM (support vector machine) applications.

4.5 Significance Tests using Probability Density Functions

In modern statistics, all classical, traditional hypothesis tests are to be avoided, as they fallaciously claim to calculate the probability of a hypothesis' validity (which, in fact, they do not). These 'classical'/outmoded tests violate Bayes' theorem; the probability of a test outcome is the probability of observing the data set, given the hypothesis and not the probability of the hypothesis, let alone the probability of the hypothesis, given the data set. Rigorous testing of significance is achieved by using probability density functions of distributions. These distributions may be the hypotheses or may be — much more generally — derived by exploratory means, such as the non-parametric kernels in KDEs (kernel density estimations).

Here I show an application, which had been used in a publication in which I am a co-author (Prossinger et al., 2023).

We use the clusters obtained after dimension reduction using an autoencoder (Chapter 3, Fig. 3-1). We use a triweight kernel for KDE. Fig. 4-3 shows a contour plot of the *pdf*s of the chosen KDE (Prossinger et al., 2023).

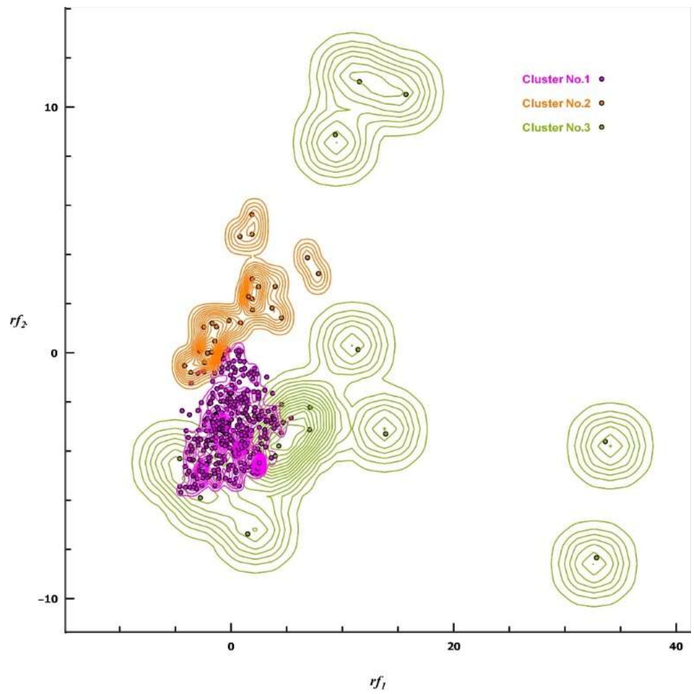


Figure 4-3 The projection of the contour plots of the likelihood surfaces obtained by the KDEs (each of the three with a triweight kernel of the functional form $\frac{35}{32}(1-u^2)^3$). Contours for each likelihood surface are in steps of $\frac{1}{15}\mathcal{L}_{\max}$ (the maximum likelihood) of the cluster). Some contours overlap.

We first show the likelihood surfaces in a 3D graph (Fig. 4-4, published in Prossinger et al., 2023).

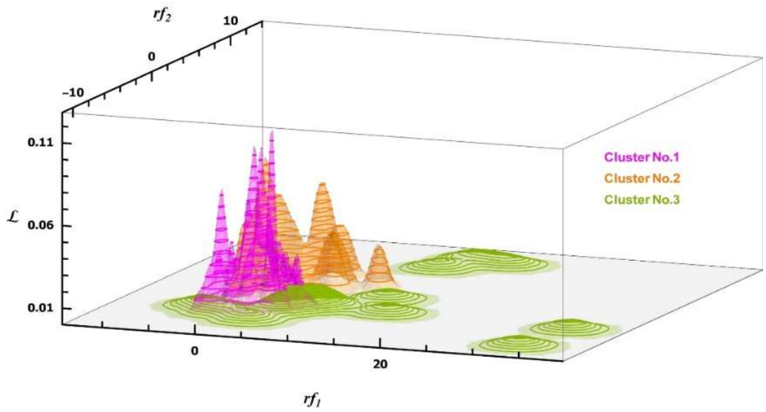


Figure 4-4 A 3D graph of the likelihood surfaces obtained by the KDEs (with the triweight kernels of the functional form $\frac{35}{32}(1-u^2)^3$). The likelihood surfaces are very broad for Cluster No.3 and very peaked for Cluster No.1. Calculation of the confusion matrix will reveal whether the overlap is significant or not.

Given the overlap of the *pdf*s of the KDEs, (as in Fig. 4-3 and Fig. 4-4) we need to determine the significance (whether the probability that the two distributions are indistinguishable is sufficiently small). This significance is expressed in a confusion matrix.

We construct the confusion matrix to estimate the significance of the overlap. For each pair of clusters (cluster_A and cluster_B), we compute the confusion matrix

$$\begin{pmatrix} \frac{pdf\left(KDE_{\text{cluster}_A}\right)>pdf\left(KDE_{\text{cluster}_B}\right)}{n_{\text{cluster}_A}} & \frac{pdf\left(KDE_{\text{cluster}_A}\right)<pdf\left(KDE_{\text{cluster}_B}\right)}{n_{\text{cluster}_A}} \\ \frac{pdf\left(KDE_{\text{cluster}_B}\right)<pdf\left(KDE_{\text{cluster}_A}\right)}{n_{\text{cluster}_B}} & \frac{pdf\left(KDE_{\text{cluster}_B}\right)>pdf\left(KDE_{\text{cluster}_A}\right)}{n_{\text{cluster}_B}} \end{pmatrix}.$$

The off-diagonal matrix entries are the probabilities of confusion (hence the name of the matrix). In this example, the overlap between Cluster #1 and Cluster #3, the confusion matrix is $\begin{pmatrix} 98.87 & 1.13 \\ 0 & 100 \end{pmatrix}$%. The overlap is insignificant and therefore the clustering algorithm did indeed partition the dimension-reduced feature vectors successfully (inferring: the clusters are truly different).

4.6 Guessing and due-to-chance Issues

In many situations involving queries, a participant makes a decision or choice. The researcher needs to consider the possibility that the participant is guessing. For example, if the query is whether a pattern is perceived as random or non-random (Boschetti et al., 2023b), the participant may not know whether it is or isn't, guesses, and by chance responded correctly. The method of estimating due-to-chance probabilities has been described in Boschetti et al. (2023b, Appendix). Here we present this Appendix *verbatim*:

"We show a method of determining whether to sets of ratings are significantly different with an example (Fig. A-1). The rating entries of the female raters for the male stimuli are n_1 and the entries for the male raters for the male stimuli are n_2; the Beta distribution is $\mathcal{Be}(n_1+1, n_2+1)$. The two boundaries $\left[\frac{1}{2}, u_{\text{upper}}\right]$ of the HDI (Highest Density Interval) are determined by $pdf\left(\frac{1}{2}\right) = pdf(u_{\text{upper}})$. (Comment: solving for u_{upper} requires computing power.) The probability of HDI is determined by

$$\text{HDI}_{\text{probability}} = \int_{\frac{1}{2}}^{u_{\text{upper}}} pdf(s)\, ds$$

Comment: this integral can be easily computed using the CDF (Cumulative distribution function) of the Beta distribution: $\text{probability} = \text{CDF}\left(\mathcal{Be}(n_1+1, n_2+1), u_{\text{upper}}\right) - \text{CDF}\left(\mathcal{Be}(n_1+1, n_2+1), \frac{1}{2}\right)$.

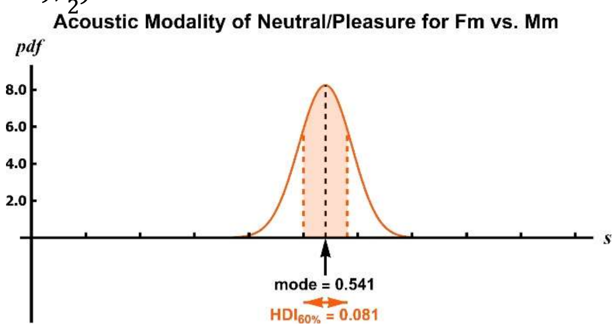


Figure A-1 A graph showing the method of determining the significance of the difference between two ratings of the same modality by raters of different sex (female versus male) of the male stimuli.

In this example, the analysis for the acoustic modality is shown, and the rating numbers of female and male stimuli are the concentration parameters of the Beta distribution (each increased by +1). s is the (Bayesian) probability. The mode of the distribution is shown, flanked by the upper and lower bounds of $HDI_{60\%}$. The dashed orange lines indicate the equal likelihoods. The area shaded in orange is the probability that the mode is significant, in this case $100\% - 60\% = 40\%$. The mode is not significantly different from $\frac{1}{2}$, and the ratings by the females and the males of the male stimuli is not significantly different at the 5% significance level.

If the computed probability is less than 95%, then the significance level is greater than 5%. In this case, the deviation of the mode from $\frac{1}{2}$ is insignificant and the differences in ratings are insignificant. (In the example shown in Fig. A-1, the probability is 60%, the significance level is therefore 40% and the observed difference in the ratings is insignificant."

Chapter 5: Visual and acoustic Stimuli

5.1 Facial expressions

To study the human ability to correctly assess two ambiguous facial expressions, namely pain and pleasure, we created our own set of stimuli, by depicting facial expressions of different emotions and affective states. As mentioned above, the elaboration and the accuracy in the rating of this type of stimuli can be affected by the spontaneity of the expressed emotion (Abramson et al., 2017; McGettigan et al., 2015). For this very reason, we avoided the development of the stimuli in laboratory conditions (by asking trained actors and actresses to mimic the facial expression in an otherwise neutral context, as we elaborated in Chapter 2). This would in principle be possible for both neutral facial expression and smile, but would border on ethical acceptability in the case of pain and pleasure. Following other authors (Fernández-Dols at al., 2011; Aviezer et al., 2012; Wenzler et al., 2016) dealing with the same issues, we decided to use audiovisual materials from which we could extrapolate via the context what type of effect was experienced and therefore expressed (Boschetti et al., 2022). Since the individual differences in facial expressivities of the expressers are crucial (but rarely discussed in scientific literature), we specifically used materials where we could find all five affective states that we wanted to test (pain, pleasure, fear, laugh, and neutral) being expressed by the same person. We could meet the condition of ethological validity by using frames from videos that depict consensual acts of extreme sexual activities (Prossinger et al., 2021a; Boschetti et al., 2021; Průšová et al., 2020; Binter et al., 2020). This type of source can be considered semi-naturalistic (Boschetti et al., 2022; Prossinger et al., 2021b): there is a clear exposition of some intense situations that elicit those responses of the expressers that we were looking for. Several such video materials were repeatedly re-viewed (literally!) by three researchers (two males, one female) to find a total of 10 videos selected, five with female expressers and five with male expressers. The selections were based upon the agreement of all three researchers regarding the context that elicited the facial expressions and their genuineness. The agreement was not based on the facial expressions (due to the ambiguous nature of the facial expressions of pain and pleasure) but because the context was sufficient for estimating the state.

The final set of stimuli consisted of 50 frames. From these frames, we extracted the facial area and the simultaneous vocalizations described in detail below. The stimuli were scaled to 600×600 pixels and the use of triangulation between the tip of the nose and pupils ensured that the proportions of the face on the screen were comparable among all stimuli. We also ensured that no background was visible within the frames presented so as to avoid contextual information (Boschetti et al., 2023c; Binter et al., 2023b; Boschetti et al., 2022; Prossinger et al., 2021b).

The final set of 50 stimuli was then pre-tested using an AI algorithm (described in detail elsewhere in this thesis), to identify the differences between the individual facial expressions, namely focusing on the pain and pleasure grimaces. The results of the pre-test study were separately published by Prossinger et al. (2022a) and confirmed the presence of significant differences between the expression of facial pain and pleasure expressions. For the rating of these stimuli we adopted the use of the valence category instead of the emotional category: to avoid biases related to the mental construct of the specific emotion and the labeling of the emotions (discussed in the Chapter 2), we offered to the raters positive, neutral and negative as possible rating options. Each of the raters was presented with the stimulus twice (Task 1 and Task 2) with the 50 stimuli in random order, to allow us to statistically estimate the consistency of their ratings. The collected data was analyzed with Dirichlet distributions. The mode has 3 components, each expressing the probability of the rating response. The closer the mode component is to 1, the higher the probability of the correct rating. (In Bayesian statistics, probability is a random continuous variable; therefore, component probabilities are along the coordinate axes.)

The results (Boschetti et al., 2022) confirmed that, while low-intensity affective states (laugh and neutral) are correctly and consistently assessed, high-intensity affective states are assessed with low accuracy: the ratings of pain and pleasure are almost equally distributed between positive and negative responses with very few neutral responses occurring. (It is important to highlight that the Dirichlet distribution allows us to compute modes lying in a triangular plane and therefore no numerical values between the positive and negative are the probability of the neutral rating.) Furthermore, in the case of high-intensity affect expressions, the ratings are observed to be consistent among raters — but they were due to chance. In other words, the participants were guessing the valence of the stimulus; surprisingly, they were doing so consistently.

5.2 Patterns

In order to study the phenomenon of illusionary visual pattern perception, we constructed our stimuli set based on randomness (Boschetti et al., 2023a; Boschetti et al., 2023b). We therefore could control for the presence/absence of a pattern. To ensure the absence of a pattern (which, in this text, we henceforth label a random pattern) in the images we used of multidimensional random number generator (specifically: an automaton) to produce maps of randomly colored squares distributed randomly within a rectangle. Below (in an overlay sense) these random colored patterns were colored, well-defined geometric figures, such as circles or pentagons. By mathematically increasing the transparency of the overlaying random maps, the patterns (geometrical figures) emerged; they became more visible with increasing transparency (Boschetti et al., 2023a). This method ensures a decrease in entropy as transparency increases and the information available to the participant increases.

We created three sets of stimuli, each composed of three images in different conditions: condition A (labeled No Pattern condition) in which the non-transparent random maps covered the pattern, a condition B (labeled Partial condition) in which the transparency of the random maps had been increased and the underlying figures were partially visible and a condition C (labeled Reveal condition) in which the transparency of the random maps is increased to a point that the geometrical figure is easily identifiable. The sequence of nine images produced this way was presented to the participants in pseudo-random order (the Reveal conditions were never presented in the beginning). The participants rated each of the stimuli using a binary option (pattern is present or pattern is absent) using keyboard keys or dedicated areas on the touch screen (Boschetti et al., 2023a).

We could differentiate between the subjects that perceived the pattern when none was present (false positive error in condition No Pattern) and the subjects that do not perceive the pattern when it was present (false negative error in Partial condition and Reveal condition). Using this methodology, we were able to study not only the phenomena of pareidolia but also its opposite: the condition in which the pattern is present but is not recognized; we call this condition apoidolia (which differs from scotomization, which is described by clinicians as a psychodynamic phenomenon).

For studying the relationship between individual differences in assessment and personality we used standardized questionnaires (translations in the Appendix). We used (1) the Coincidence Questionnaire (CQ; Bressan, 2002) to measure the experienced perceived coincidence; (2) the Religion Commitment Inventory (RCI; Worthington et al., 2003) to measure differences in the strength of religious beliefs; (3) the Illusory Beliefs Inventory (IBI; Kingdon et al., 2012) to measure magical beliefs and spirituality and (4) the Rational/Experiential Multimodal Inventory (REIm; Norris & Epstein, 2011) to measure thinking styles.

We analyzed the Dirichlet distributions of the response distributions of each query (or item) separately for participants that made a perceptual error in rating the stimuli (either false positive or false negative based on the condition) and for participants that correctly identified the stimulus. To test whether these distributions were significantly different between participants who correctly identified the absence/presence of the pattern and pareidolia/apoidolia participants, we constructed the confusion matrix of the Dirichlet distribution. The details of this methodology are described in Chapter 3, Section 3.5.

This type of analysis, clearly showed that, of all the psychological variables that we tested with the questionnaires, the thinking style (Norris & Epstein, 2011) was significantly different. Above 70% of the queries of each subscale (Rationality, Intuition, Imagination, and Emotionality) have different Dirichlet distributions in participants that exhibited pareidolia or apoidolia from the participants that correctly rated the stimulus. In a further study (Boschetti et al, 2023b) we replaced the colored stimuli with black and white stimuli (similar to QR codes) thereby maximizing the visual contrast in the elements of the stimulus. Also in this case we were able to statistically control for the presence or absence of patterns in the images. To minimize guessing, we presented each stimulus four times, with one spatial rotation and two axis reflections so as to avoid any learning effect. Guessing in this context is known as *The Lady Tasting Tea Problem* (Fisher, 1956): if a two-option challenge is presented to a respondent, the chance of correct identification is $\frac{1}{2}$; this chance is far too high; if the challenge is presented four times in random order, then chance of correct identification by guessing is $\frac{1}{2^4} = \frac{1}{16} \sim 6\%$. We did not repeat the presentations five times (then the chance of guessing would be $\frac{1}{2^5} = \frac{1}{32} \sim 3\%$ — very much lower than the conventional 5% significance level), because of possible uncontrollable effects due to the tiring of the participants. This is an example of a classical trade-off between two effects: achieving a low significance level versus the inability to maintain consistency in experimental repeats (as participants get tired, they behave differently from when they were fresh). We used an Artificial Neural Network (an autoencoder) to dimension-reduce the feature vector of the responses and then identified clusters (details described in Chapter 3, Section 3.3). We found that the raters based on their responses on thinking style (specifically the REIm subscales for Intuition and for Rationality, separately) and their results of the pattern identification task. We identified two clusters for each subscale, confirming the association between identification patterns and thinking style (Boschetti et al., 2023b). From these results, it emerges that what is relevant in association with the pattern perception is not the specific type of belief but the thinking style (i.e. experiential vs rational), which may function as a substrate for a specific type of belief.

5.3 Auditory stimuli: vocalizations

As mentioned above, not only facial expressions, but also vocalizations of affective states were presented; we anticipated it to be subject to the *emotion intensity paradox* (Holz et al., 2021): the more intense the vocalizations, the more difficult it is to correctly assess the valence. From a theoretical perspective, the acoustic stimulus may function primarily to gain attention; this has been previously found when comparing screams (intense emotional vocalizations) with regular speech in the context of accuracy and rapidity of localization (Arnal et al., 2015). Since the intensity of the experience is very high, we can assume that the situation that evokes the vocalization is rich in contextual information about the valence (positive or negative), the aim of the vocalizations would be to capture and orient the attention of those observing (i.e. the receivers) towards the expresser and towards the context.

To investigate the assessment of vocalization of affective states, we used the same methodology as for the study of facial expressions (Boschetti et al., 2022; Binter et al., 2023a) the raters assessed the valence (positive, neutral, and negative) of 50 stimuli (10 for each vocalization: laugh, neutral, fear, pain, pleasure) twice (Task 1 and Task 2). The possible response options for the ratings were positive, neutral and negative. We could not pre-test the stimuli in this case because the algorithms for such tasks have yet to be developed. However, the vocalizations were simultaneous with the pre-tested facial expressions. Our assumption is that, just as the facial expression are distinguishable, in the same way the simultaneous vocalizations resemble the state the expresser finds himself/herself in. For the accuracy of the assessment, we obtained similar results as for the facial expressions: while vocalization of affective states with low intensity where rated with above 85% of correct responses, the high-arousal affective states were not correctly identified. In contrast to the results we obtained for facial expressions, the due-to-chance probability for none of the ratings for vocalizations was due to chance (Binter et al., 2023a). This infers that the raters were convinced of their assessment (they not guessing) even when the assessment was wrong.

Furthermore, we tested the consistency of the ratings (between Task 1 and Task 2). Surprisingly participants were not consistent in their rating. Taken together, these results showed a very interesting outcome: even if the raters were rating the high arousal vocalizations with low assessment accuracy, they were not guessing, but in this case they were not consistent in their ratings either. They were making a different mistake when providing a rating during Task 2 (this possibility is due to the fact that they could choose among three options — negative, neutral, and positive), and in both tasks cases they were convinced of the option they chose.

5.4 Multimodal: visual and auditory stimuli

To investigate how the assessment of ambiguous displays can benefit from multimodal redundancy of information and also whether they are affected by sensory cross-talk, we presented our raters the facial expression together with the vocalizations of the affective states (Boschetti et al., 2023c; Binter et al., 2023a).

As had been mentioned previously, when raters assessed the valence (positive, negative, or neutral) of the visual stimulus (facial expression) and the auditory stimulus (the vocalizations) separately, they were not accurate and, in the case of facial expressions, the results were due to chance while in the case of vocalizations, the raters were not guessing.

To test the effect of redundancy in assessment accuracy in the congruent condition (pleasure visually and acoustically, say) we constructed multimodal stimuli in which the visual and auditory signals are congruent with each other and, therefore, they carry redundant information regarding the same affective states (Binter et al., 2023a). We presented the bimodal stimuli of two affective states with low intensity (neutral and smile/laugh) and two with high intensity (pain and pleasure). Surprisingly we did not find any improvement in the correctness of the assessment.

Neutral and smile/laugh were rated with very high accuracy, and pain and pleasure with very low accuracy; these are the same results we found for the facial expression and vocalization when presented separately. However, the raters' convictions and their beliefs in their assessments changed: compared to the situation when the participants were presented only with the visual stimuli, the results were not due to chance (for two specific cases: females rating female pleasure stimuli and males rating male pleasure stimuli; in these cases, the ratings were due to guessing). In the cases of congruent modality, the redundancy of information carried by different modalities did not improve the accuracy of the ratings, but it did change the level of confidence that the raters had in their ratings. Generally, they were more confident than when rating only the facial expression but less confident than when rating only the vocalizations (Binter et al., 2023a). This should be taken into account when complex stimuli are constructed.

To investigate the cross-talk between modalities of ambiguous stimuli, we constructed incongruent multimodal stimuli (Boschetti et al., 2023a) in which the facial expressions of three affective states (pain, pleasure, and neutral) were paired with unmatched vocalizations (i.e. the facial expression of pleasure with the vocalization of pain); there are six such pairings. These types of stimuli do not occur naturally; they therefore lack ecological validity. Nonetheless, they can bring insight into how the signal carried by one sensory modality interferes with the signal carried by the other modality when both are experienced simultaneously. Also, it is important to consider the amount of audio-visual media the individuals are exposed to every day, where such options are certainly available (and used).

It is important to point out that it is impossible for a rater to rate *both* presented modalities correctly (since the visual and acoustic stimulus carried information from contradicting affective states). However, in the case of cross-talk, could rate both *incorrectly* (because a third choice existed: if, for example, the correct visual stimulus was negative and the correct acoustic stimulus was neutral, the rater could still choose the response option neutral).

In our results show that the acoustic modality is — again — more reliable and, with the exception of two incongruent combinations, the information gathered through this modality suppressed the visual information (Boschetti et al., 2023c). These results are not surprising if we consider that vocalization, compared to facial expression, is not rated due to chance. This conviction in the assessment of vocalization may manifest itself for incongruent bimodal stimuli by suppressing the information from the visual signal. The only case in which we observed the converse is when the visual stimulus was depicting pleasure and was combined with the vocalization of pain; in this case, the stimulus was rated as positive (indicating that the visual stimulus overrode the acoustic one).

In disagreement with the published studies, we found evidence of sensory cross-talk only in one case: when the visual stimulus depicting a neutral expression is paired with a vocalization of pleasure. In this case, none of the signals (visual or auditory) were suppressed but the two interfered with each other; the resulting rating did not match either of the modalities.

5.5 Embodied stimulation: effect of the arousal of the rater

A factor that can influence the perception of the stimuli is the inner state of the rater. This factor is specifically important when highly arousing stimuli are studied. In real life situations, the subject rarely stays calm when encountering highly arousing situations (i.e. winning or losing a potential sexual interaction) and this is in marked contrast to common rating assessment tasks in a research context (often conducted in a laboratory).

These situations may involve emotional coupling and affect mirroring and cause a dynamic attribution process (Hasson & Frith, 2016). Indeed, the state of the raters was previously shown to affect the perception of the expresser, especially concerning the assessment of arousal and valence (Pell, 2005). The concept of arousal is functionally related to the sympathetic nervous system (Dawson et al., 2000), which is responsible for mobilizing the organism's resources to meet internal physiological demands as well as those of the external environment (Salvia et al., 2012). To manipulate this variable, in previous studies the stress was induced by exposure to heights and situations wherein participants expected an electrical shock.

A modern, more ethical, yet equally reliable, alternative to increase physiological arousal is the Cold Pressor Task (CPT; Bullinger et al., 1984; Binter et al., 2022a). It consists of immersing a subject's extremity (usually the foot) into ice water for a specified period of time (Mitchell et al., 2004). We adopted this procedure to increase the arousal in one group of raters (labeled the experimental group); their lower right limb (the crus) was immersed in cold water (2–4 °C) for 1½ minutes (Binter et al., 2023a; Boschetti et al., 2022; Binter et al., 2022a). In contrast, the control group's lower right leg was immersed in water at room temperature. After this procedure, the raters took part in the same process of assessment of facial expressions and vocalizations as previously described. We did not repeat the ratings twice because the CPT impact on the cortisol release is most effective in the first 20 minutes after the immersion and repeating the ratings scenario would take longer than that.

For facial expressions (Boschetti et al., 2022), our results show a significant increase in accuracy (a shift towards a more positive rating) in the case of laughter and pleasure but only for images depicting male expressers. For vocalizations (Binter et al., 2023a), we found a very different result: the probability of correct attribution significantly decreased in the stressed group for laugh and for neutral but also in this case only for male expressers' vocalizations.

These results show that there is a significant effect of increasing the raters' arousal in the assessment but this effect is not generalizable. It did not affect the assessment for all of the stimuli (it depended on the sex of the expresser and the specific affective state) and it affected the visual and auditory modality differently. The results we have published open the possibility of proposing novel hypotheses, which can then be studied in suitably designed test situations. We can assume that a better description of the mechanism will become possible.

Chapter 6: Implications for Applications

There are potential applications for algorithms employing feature vector extraction: as shown in this thesis, these algorithms can distinguish between the facial expressions. One application: they can be applied to both online videos and to surveillance camera videos with the intent to spot facial expressions of those in danger or injured or otherwise at risk of harm.

The findings related to the assessment of facial expressions and vocalizations of pain and pleasure have two different, yet parallel, implications based on their source — the real world on the one hand and audiovisual media on the other.

While interacting with others in the real-world, the information gathered from facial and vocal displays is obviously insufficient for making an appropriate decision. Another application: specifically, in case of sexual intercourse, the correct assessment of pain and pleasure is — arguably — very important. In this situation, the contextual cues that elicit the affective state are insufficient to distinguish between the two: the same activity can indeed elicit both. Furthermore, other nonverbal behaviors, such as body posture, could serve as additional cues to disambiguate the facial/vocal displays. Arguably, the person's perceptions are limited to facial expression and the vocalization. Therefore, rarely are other nonverbal cues taken into consideration when assessing the partner's affective state. Verbal communication concerning the experienced affective states should then be preferred in moments of doubt. Additionally, the current evaluative methods used by sexologist to assess sexual aggressiveness are mainly focused on the visual stimulation, while it would be beneficial to also include evaluations based on vocalizations.

Third application: how pattern perception relates to individual differences in rationality and intuition brings novel insight into the study of individuals' perceptions. Pattern perception tests can also be applied as an uncomplicated and nonetheless interesting method to uncover such personality traits. Doubtless, further research is to be conducted.

Chapter 7: Conclusion

Humans are sometimes referred to as pattern-seeking animals. Based on our perception of patterns in our environment, we can extract information, help contribute to our survival and usually benefit from our actions. The decisions we make every day are based on innate mechanisms, development stages, long-term experiences, personality, individual differences, and momentary disposition (emotional, physiological, mental, etc.).

Rather than using pre-tested database stimuli, all stimuli I used were generated either by computer graphics tools or by extracting facial expressions and vocalizations from freely available online videos. In both cases, I could avoid criticisms regarding lack of ecological validity. Our mutual interaction with others in society depends on our ability to assess their affective state, more often than not based on their affective displays. Specifically, it is important to be able to assess the highly-intensive ones. Research suggests that the accuracy of assessment is low, due to *intensity paradox*. Pre-testing using human raters is therefore in vain. As a consequence, I applied novel feature extracting algorithms that are more reliable for these types of stimuli than conventionally used FACS models.

As expressly pointed out in the title of this thesis, it addresses the issue of uncertainty entailed in human perception. I discovered that the use of objectively generated patterns permits a method of quantifying (with AI and clustering algorithms) the magnitude of uncertainty. Future investigations can now focus how the interplay of signal corruption and the intensity paradox leads to heretofore unexpected insights — whatever they will be.

The stimuli were presented to a large, international sample of participants who rated the stimuli repeatedly using a categorical rating procedure that is simple and reliable and avoids all the fallacious pitfalls of Likert-type-scale uses that had already been criticized by previous researchers.

By using a Bayesian statistical approach, specifically Beta and Dirichlet distributions, to derive the off-diagonal elements of confusion matrix (in order to determine significance), I was able to prove the influence of thinking style on visual pattern perception. Also, while studying the influence of the intensity of the affective state of the expresser on the ability to assess the valence of the expression by the rater, I was able to show that ambiguous facial expressions, and accompanying vocalizations, of high-intensity affects (pain, pleasure) were rated with a low probability of correct response. Furthermore, through the application of due-to-chance analysis, I could determine the probability that the participants were guessing: for the visual stimuli this probability was high, as opposed to the auditory stimuli, in which case the probability was low.

Application of the Cold Pressor Task allowed me to study the shifts of ratings of the facial expression and the vocalizations when the raters themselves are in a high arousal condition. Interestingly, all shifts occurred only for male stimuli.

I found that, in the congruent modalities condition, the probability of correct ratings of the highly intensive displays did not increase. But only when males assessed female stimuli and females assessed male stimuli of pleasure were the ratings due to chance.

In the incongruent condition (in which no correct rating can exist) I found that, for most of the pairings, the ratings of the acoustic modality predominated over the visual one — which is a further, perhaps unexpected, contribution to current scientific knowledge.

Several novelties were included in my research: the greater rigor of stimuli preparation, state-of-the-art statistics of rating procedures evaluations, and outcome interpretations involving significance tests using confusion matrices. In my thesis, I opted for a multidisciplinary approach connecting evolutionary, biological, psychological, and ethological approaches with those of the physical sciences and of mathematical statistics (not only, but also involving the use of Artificial Intelligence).

I expect that these outcomes will contribute to future researchers adopting these methods (both theoretical and experimental) in their scientific practice.

References

Abramson, L., Marom, I., Petranker, R., & Aviezer, H. (2017). Is fear in your head? A comparison of instructed and real-life expressions of emotion in the face and body. *Emotion*, 17(3), 557

Akçay, Ç., & Beecher, M. D. (2019). Multi-modal communication: song sparrows increase signal redundancy in noise. *Biology Letters*, 15(10), 20190513.

Aristotele (4th Century B.C.). *Collected Works.* In Laërtius, D. (3rd Century A.D.). *Lives and Opinion of Eminent Philosophers*. Rome.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology*, 25(15), 2051–22056.

Aviezer, H., Messinger, D. S., Zangvil, S., Mattson, W. I., Gangi, D. N., & Todorov, A. (2015). Thrill of victory or agony of defeat? Perceivers fail to utilize information in facial movements. *Emotion*, 15(6), 791.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, 338(6111), 1225–1229. doi: 10.1126/science.1224313

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., ... & Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychol Sci*, 19(7), 724-732.

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in cognitive sciences*, 4(1), 29-34.

Binter, J., Boschetti, S., Hladký, T., Prossinger, H., Wells, T. J., Jílková, J., & Říha, D. (2022a). Quantifying the Rating Performance of Ambiguous and Unambiguous Facial Expression Perceptions Under Conditions of Stress by Using Wearable Sensors. *In International Conference on Human-Computer Interaction* (pp. 519-529). Springer, Cham. doi: 10.1007/978-3-031-17615-9_36

Binter, J., Boschetti, S., Hladký, T. & Prossinger, H. (2023a, *submitted*) Ouch!" or "Aah!": Are Vocalizations of 'Laugh', 'Neutral', 'Fear', 'Pain' or 'Pleasure' Reliably Rated? *Human Ethology*.

Binter, J., Prossinger, H., Hladký, T., & Boschetti, S. (2023b, *submitted*). " Wow!" and" Aah": Bimodal Stimuli of Affective States with High and Low Intensity Displays.

Scientific Reports.

Blakeslee, S. (2006). Cells that read minds. *New York Times*, 10(1).

Blalock Jr, H. M. (1960). *Social statistics*.

Boschetti, S., Hladký, T., Říha D. & Binter, J. (2021). Sex a jídlo – slasti a cyklické stavy, které procházejí obdobnými procesy: jídlo kořeněné chilli a sex kořeněný sadomasochismem mají společné rysy. Sexuológia (Sexology), 1, 31 – 38.

Boschetti, S., Prossinger, H., Hladký, T., Machová, K., & Binter, J. (2022). "Eye can't see the difference": Facial Expressions of Pain, Pleasure, and Fear Are Consistently Rated Due to Chance. *Human Ethology*, Volume 37, 046-072. doi: 10.22330/he/37/046-072

Boschetti, S., Binter, J. and Prossinger, H. (2023a, *in press*) Apoidolia: A New Psychological Phenomenon Detected by Pattern Creation with Image Processing Software together with Dirichlet Distributions and Confusion Matrices. *Proceedings in Adaptation, Learning and Optimization.*

Boschetti, S., Prossinger, H., Hladký, T., Říha, D., Příplatová, L., Kopecký, R. & Binter, J. (2023b, *in press*). Are Patterns Game for Our Brain? AI identifies individual Differences in Rationality and Intuition Characteristics of Respondents Attempting to Identify Random and Non-random Patterns. *In International Conference on Human-Computer Interaction.*

Boschetti, S., Prossinger, H., Prossinger-Beck, V., Hladký, T., Říha, D. & Binter, J. (2023c, *in press)*. Never correct: The novel analysis of differing visual (facial expression) and acoustic (vocalization) bimodal displays of the affective states "pain", "pleasure", and "neutral". *In International Conference on Human-Computer Interaction*.

Bressan, P. (2002). The connection between random sequences, everyday coincidences, and belief in the paranormal. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition,* 16(1), 17-34.

Bullinger, M., Naber, D., Pickar, D., Cohen, R. M., Kalin, N. H., Pert, A., & Bunney Jr, W. E. (1984). Endocrine effects of the cold pressor test: relationships to subjective pain appraisal and coping. *Psychiatry Research*, 12(3), 227–233. doi: 10.1016/0165-1781(84)90028-3

Byosiere, S. E., Chouinard, P. A., Howell, T. J., & Bennett, P. C. (2018). What do dogs (Canis familiaris) see? A review of vision in dogs and implications for cognition research. *Psychonomic bulletin & review*, 25, 1798-1813.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in* *cognitive sciences*, 11(12), 535-543.

Chen, C., Crivelli, C., Garrod, O. G., Schyns, P. G., Fernández-Dols, J. M., & Jack, R. E. (2018). Distinct facial expressions represent pain and pleasure across cultures. *Proceedings of the National Academy of Sciences*, 115(43), E10013–E10021. doi: 10.1073/pnas.1807862115

Corvin, S., Fauchon, C., Peyron, R., Reby, D., & Mathevon, N. (2022). Adults learn to identify pain in babies' cries. *Current Biology*, 32(15), R824-R825.

Cronin, T. W., Bok, M. J., Marshall, N. J., & Caldwell, R. L. (2014). Filtering and polychromatic vision in mantis shrimps: themes in visible and ultraviolet vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1636), 20130032.

Darwin, C. (1872). *The expression of the emotions in man and animals*, NY: D. Appleton and Company

Dawson, M., Schell, A., Filion, D. (2000). The electrodermal system. *Handbook of Psychophysiology*. Cambridge University Press

Deneve, S., & Jardri, R. (2016). Circular inference: mistaken belief, misplaced trust. *Current Opinion in Behavioral Sciences*, 11, 40-48.

Dolan, R. J., Morris, J. S. & deGelder, B. (2001) Crossmodal binding of fear in voice and face. *Proc Natl Acad Sci* U S A 98, 10 006– 10 010.

Ekman, P. (2006). *Darwin and facial expression: A century of research in review*. NY: Academic Press.

Ekman P, Friesen W. (1978). *Facial Action Coding System: investigator's guide*. Palo Alto, CA: Consulting Psychologists Press

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin*, 128(2), 203.

Fairhurst, M. (2014). Not 5 but 33 senses. *Arts & Humanities Research Council*.

Fernández-Dols, J. M., Carrera, P., & Crivelli, C. (2011). Facial behavior while experiencing sexual excitement. *Journal of Nonverbal Behavior*, 35(1), 63-71.

Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother-stranger face discrimination by the newborn. *Infant Behavior and development*, 7(1), 19-25.

Fisher, R. A. (1956). Mathematics of a lady tasting tea. *The world of mathematics*, 3(part 8), 1514-1521.

Grammer, K., Kruck, K.B. & Magnusson, M.S. (1998). The courtship dance: patterns of nonverbal synchronization in opposite-sex encounters. *J Nonverbal Behav*, 22, 3.

Grammer, K. (1990). Strangers meet: laughter and nonverbal signs of interest in opposite-sex encounters. *J Nonverbal Behav*, 14, 209.

Gregory, R. L. (1997). Knowledge in perception and illusion. Philosophical Transactions of the Royal Society of London. *Series B: Biological Sciences*, 352(1358), 1121-1127.

Hamm, J., Kohler, C. G., Gur, R. C., & Verma, R. (2011). Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of neuroscience methods*, 200(2), 237-256.

Haselton, M., & Galperin, A. (2012). Error management and the evolution of cognitive bias. *Soc. Think. Interpers. Behav*. 45, 63.

Hasson, U., & Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1693), 20150366. doi: 10.1098/rstb.2015.0366

Hjortsjö, C. H. (1970). *Man's face and mimic language*. Lund, Sweden: Studentlitteratur.

Holz, N., Larrouy-Maestri, P., & Poeppel, D. (2021). The paradoxical role of emotional intensity in the perception of vocal affect. *Scientific reports*, 11(1), 1–10.

Hughes, S. M., & Nicholson, S. E. (2008). Sex differences in the assessment of pain versus sexual pleasure facial expressions. *Journal of Social, Evolutionary, and Cultural Psychology*, 2(4), 289.

Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research. *Psychological Bulletin*, 115(2), 288–299. doi: 10.1037/0033-2909.115.2.288

Jacobs, G. H., Deegan, J. F., Crognale, M. A., & Fenwick, J. A. (1993). Photopigments of dogs and foxes and their implications for canid vision. *Visual Neuroscience*, 10(1), 173-180.

Johnson, D. D. (2009). The error of God: Error management theory, religion, and the evolution of cooperation. *In Games, groups, and the global good* (pp. 169-180). Springer, Berlin, Heidelberg

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476), 2109-2128.

Kingdon, B. L., Egan, S. J., & Rees, C. S. (2012). The Illusory Beliefs Inventory: A new measure of magical thinking and its relationship with obsessive compulsive disorder. *Behavioural and Cognitive Psychotherapy* 40.1, 39–53.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in methods and practices in psychological science*, 1(2), 270-280.

Leongómez, J., Binter, J., Kubicová, L., et al. (2014). Vocal modulation during courtship increases proceptivity even in native speakers. *Evol Hum Behav*, 35(6), 489-496

Liu, R., Hannum, M., & Simons, C. T. (2019). Using immersive technologies to explore the effects of congruent and incongruent contextual cues on context recall, product evaluation time, and preference and liking during consumer hedonic testing. *Food Research International*, 117, 19-29.

Lorenz, K. (1939). Vergleichende Verhaltensforschung. *Verhandlungen der Deutschen Zoologischen Gesellschaft Zoologischer Anzeiger*, Supplementband, 12, 69–102.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). Karolinska directed emotional faces. *Cognition and Emotion*.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior research methods*, 47, 1122-1135.

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., & Scott, S. K. (2015). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral cortex*, 25(1), 246-257.

Metaxakis A, Petratou D, Tavernarakis (2018). Multimodal sensory processing in Caenorhabditis elegans. *Open Biol*. 8: 180049. http://dx.doi.org/10.1098/rsob.180049

Mitchell, L. A., MacDonald, R. A., & Brodie, E. E. (2004). Temperature and the cold pressor test. *Journal of Pain*, 5(4), 233–237. doi: 10.1016/j.jpain.2004.03.004

Müller, V. I., Habel, U., Derntl, B., et al. (2011). Incongruence effects in crossmodal emotional integration. *Neuroimage* 54, 2257–2266.

Norris, P., & Epstein, S. (2011). An experiential thinking style: Its facets and relations with objective and subjective criterion measures. *Journal of personality*, 79(5), 1043-1080.

Pell, M. D. (2005). Nonverbal emotion priming: evidence from the ‘facial affect decision task’. *Journal of Nonverbal Behavior*, 29(1), 45–73. doi: 10.1007/s10919-004-0889-8

Philips, S. M., & Pitton, J. W. (2008). Similarity-Based Perceptual Feature Identification for

Active Sonar Signal Classification. *Journal of the Acoustical Society of America*, 123(5), 3344.p. 387

Pollak, S. D., Messner, M., Kistler, D. J., & Cohn, J. F. (2009). Development of perceptual expertise in emotion recognition. *Cognition*, 110(2), 242-247.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev Psychopathol*, 17(3), 715-734.

Prossinger, H., Boschetti, S., Hladký, T., Říha D. & Binter, J. (2021a). Konzumace pornografie obyvatel ČR a její vztah k demografickým proměnným: klastrování pomocí umělé inteligence. *Sexuológia (Sexology)*, 1, 7 - 16.

Prossinger, H., Hladky, T., Binter, J., Boschetti, S., & Riha, D. (2021b). Visual Analysis of Emotions Using AI Image-Processing Software: Possible Male/Female Differences between the Emotion Pairs "Neutral"–"Fear" and "Pleasure"–"Pain". *In The 14th PErvasive Technologies Related to Assistive Environments Conference* (pp. 342-346).

Prossinger, H., Hladký, T., Boschetti, S., Říha, D., & Binter, J. (2022a). Determination of "Neutral"–"Pain","Neutral"- "Pleasure", and "Pleasure"–"Pain" Affective State Distances by Using AI Image Analysis of Facial Expressions. *Technologies*, 10(4), 75. doi: 10.3390/technologies10040075

Prossinger, H., Boschetti, S., Říha, D., Pacovský, L. & Binter, J. (2023, *in press*) From Stone Age to New Age Statistics: How Neural Networks Overcome the Irreproducibility Problems in Choice Based Profile Creation. *In International Conference on Human-Computer Interaction*.

Průšová, D., Boschetti, S., Hladký, T., Říha, D., Wells, T.J. & Binter, J. (2020). Vliv pornografie na přejímání emocí a sexuálního chování. *Sexuológia (Sexology)*, 2, 31 – 35.

Raine, J., Pisanski, K., & Reby, D. (2017). Tennis grunts communicate acoustic cues to sex and contest outcome. *Animal Behaviour*, 130, 47-55.

Roberson, D., Damjanovic, L., & Kikutani, M. (2010). Show and tell: The role of language in

categorizing facial expression of emotion. *Emotion Review*, 2(3), 255–260. doi: 10.1177/1754073910361979

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161. doi: 10.1037/h0077714

Salvia, E., Guillot, A., & Collet, C. (2012). Autonomic nervous system correlates to readiness state and negative outcome during visual discrimination tasks. *International Journal of Psychophysiology*, 84(2), 211–218. doi: 10.1016/j.ijpsycho.2012.02.010

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *Int J Psychol Psychoanal*, 46(6), 401- 435.

Sebe, N., Lew, M. S., Sun, Y., Cohen, I., Gevers, T., & Huang, T. S. (2007). Authentic facial expression analysis. *Image and Vision Computing*, 25(12), 1856-1863.

Smith, R. C., & Baker, K. S. (1981). Optical properties of the clearest natural waters (200–800 nm). *Applied optics*, 20(2), 177-184.

Sourina, O. & Liu, Y. (2011) A fractal-based algorithm of emotion recognition from EEG using arousal-valence model. *Biosignals*. Rome, Italy, 209-214

Temple, S. E., How, M. J., Powell, S. B., Gruev, V., Marshall, N. J., & Roberts, N. W. (2021). Thresholds of polarization vision in octopuses. *Journal of Experimental Biology*, 224(7), jeb240812.

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annu. Rev. Neurosci.*, 31, 411-437.

Van Der Tempel, J., & Alcock, J. E. (2015). Relationships between conspiracy mentality, hyperactive agency detection, and schizotypy: Supernatural forces at work?. *Personality and Individual Differences*, 82, 136-141.

Van Der Zant, T., & Nelson, N. L. (2021). Motion increases recognition of naturalistic postures but not facial expressions. *Journal of Nonverbal Behavior*, 45(4), 587–600. doi: 10.1007/s10919-021-00372-4

von Uexküll J. (1926). *Theoretical Biology*. New York, NY: Harcourt, Brace & Company, Inc.

Wada Y., Kitagawa N., Noguchi K. (2003). Audio-visual integration in temporal perception. *Int J Psychophysiol* 50(1–2) 117–124. 10.1016/S0167–8760(03)00128–4

Walker, A. C., Turpin, M. H., Stolz, J. A., Fugelsang, J. A., & Koehler, D. J. (2019). Finding meaning in the clouds: Illusory pattern perception predicts receptivity to pseudo-profound bullshit. *Judgment and Decision Making*, 14(2), 109-119.

Waller, B. M., Cray Jr, J. J., & Burrows, A. M. (2008). Selection for universal facial emotion. *Emotion*, 8(3), 435.

Wenzler, S., Levine, S., van Dick, R., Oertel-Knöchel, V., & Aviezer, H. (2016). Beyond pleasure and pain: Facial expression ambiguity in adults and children during intense situations. *Emotion*, 16(6), 807. doi: 10.1037/emo0000185

Willard, A. K., & Norenzayan, A. (2013). Cognitive biases explain religious belief, paranormal belief, and belief in life's purpose. *Cognition*, 129(2), 379-391.

Williams, D. R., & MacLeod, D. I. (1979). Interchangeable backgrounds for cone afterimages. *Vision Research*, 19(8), 867-877.

Wittfoth, M., Schröder, C., Schardt, D. M., Dengler, R., Heinze, H. J. & Kotz, S. A. (2010). On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. *Cereb Cortex*. 2010 20(2):383-92.

Worthington Jr, E. L., Wade, N. G., Hight, T. L., Ripley, J. S., McCullough, M. E., Berry, J. W., ... & O'Connor, L. (2003). The Religious Commitment Inventory-10: Development, refinement, and validation of a brief scale for research and counseling. *Journal of counseling psychology*, 50(1), 84.

Yokoi, K., Nishio, Y., Uchiyama, M., Shimomura, T., Iizuka, O., & Mori, E. (2014). Hallucinators find meaning in noises: pareidolic illusions in dementia with Lewy bodies. *Neuropsychologia*, 56, 245-254.

Zhang, F., & Parmley, M. (2015). Emotion attention and recognition of facial expressions among close friends and casual acquaintances. *Journal of Social and Personal Relationships*, 32(5), 633-649.

Zhou, L. F., & Meng, M. (2020). Do you see the "face"? Individual differences in face pareidolia. *Journal of Pacific Rim Psychology*, 14, e2.

Appendix: Translations of Questionnaires

I would like to thank for the collaboration in the translation of the questionaires: Martin Hůla, Robin Kopecký, Jelena Příplatová, Jakub Binter, Tomas Hladký, Filippo Talia and Lorenzo Bazzoni.

Italian translation of **Religious Commitment Inventory-10**

Worthington Jr, E. L., Wade, N. G., Hight, T. L., Ripley, J. S., McCullough, M. E., Berry, J. W., & O'Connor, L. (2003). The Religious Commitment Inventory-10: Development, refinement, and validation of a brief scale for research and counseling. *Journal of counseling psychology*, 50(1), 84.

Ogni voce è valutata come segue: 1 = non è assolutamente vero per me, 2 = è un po' vero per me, 3 = moderatamente vero per me, 4 = per lo più vero per me, o 5 = totalmente vero per me.

1. Leggo spesso libri e riviste riguardo la mia fede.
2. Faccio delle offerte economiche alla mia organizzazione religiosa.
3. Investo del tempo cercando di comprendere più a fondo la mia fede.
4. La religione è particolarmente importante per me perché risponde a diverse domande riguardo al significato della vita.
5. Le mie credenze religiose sono ciò che sta dietro al mio approccio alla vita.
6. Mi piace passare del tempo con altri membri della mia organizzazione religiosa.
7. Le mie credenze religiose influenzano tutti i rapporti della mia vita.
8. E' importante per me passare dei periodi di tempo in riflessioni religiose private.
9. Mi piace lavorare nelle attività della mia comunità religiose.
10. Mi mantengo ben informato sul mio gruppo religioso locale e ho una certa influenza nelle sue decisioni.

Czech translation of **Religious Commitment Inventory-10**

Worthington Jr, E. L., Wade, N. G., Hight, T. L., Ripley, J. S., McCullough, M. E., Berry, J. W., & O'Connor, L. (2003). The Religious Commitment Inventory-10: Development, refinement, and validation of a brief scale for research and counseling. *Journal of counseling psychology*, 50(1), 84.

Každá otázka je hodnocena na následující škále: 1 = vůbec pro mě neplatí, 2 = trochu pro mě platí, 3 = středně pro mě platí, 4 = spíše pro mě platí, 5 = naprosto pro mě platí.

1. Často čtu knihy a časopisy, které se zabývají mou vírou.
2. Finančně přispívám na fungování své církevní organizace.
3. Věnuji čas snahám o prohloubení porozumění mé víře.
4. Náboženství je pro mne velmi důležité, protože dává odpověď na mnoho otázek týkajících se smyslu života.
5. Náboženské postoje určují můj celkový přístup k životu.
6. Rád/a trávím čas slidmi, se kterými sdílím svou víru.
7. Má víra ovlivňuje všechna máživotní rozhodnutí.
8. Je pro mne důležité trávit chvíle vlastním rozjímáním a reflektováním náboženských myšlenek.
9. Rád/a se podílím na aktivitách své náboženské organizace.
10. Vím hodně o své místní náboženské komunitě a mám určitý vliv na rozhodnutí, která dělá.

Italian translation of **Questionnaire on Coincidence**

Bressan, P. (2002). The connection between random sequences, everyday coincidences, and belief in the paranormal. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 16(1), 17-34.

In generale quanto spesso ti capita di imbatterti in curiose o significative coincidenze?

Per favore indicare in una scala da 1 a 5, dove 1=mai, 2=una o due volte, 3=qualche volta, 4=molte volte, 5=molto spesso

1) Ci sono molti tipi di coincidenze. Quanto spesso ti è capitato di vivere una coincidenza in ognuna delle seguenti categorie?

(a) Una serie di nomi, numeri o eventi dello stesso tipo (come incorrere ripetutamente in poche ore in una parola mai sentita prima) *[1,2,3,4,5]*

(b) Associazioni spontanee (come pensare a una persona e incontrarla inaspettatamente poco dopo) *[1,2,3,4,5]*

(c) "Il mondo è piccolo" (come incontrare una persona che non si vede da molto tempo in un posto improbabile) *[1,2,3,4,5]*

(d) Percezione di qualcosa di fisicamente lontano (come preoccuparsi per una persona nell'esatto momento in cui sta avendo un incidente) *[1,2,3,4,5]*

(e) Percezione di qualcosa distante nel tempo (come fare un sogno che poi diventa realtà) *[1,2,3,4,5]*

(f) Soluzione inaspettata di un problema (come incontrare un amico che vuole vendere il suo computer quando tu ne stai cercando uno) *[1,2,3,4,5]*

(g) "Angelo custode" (come non arrivare in tempo a un colloquio per poi scoprire che è stato meglio così, perché si ha un'occasione migliore che altrimenti si sarebbe persa) *[1,2,3,4,5]*

2) Pensi che le coincidenze siano dovute a:

(a) Puro caso *[Sì, no, non so]*

(b) Destino *[Sì, no, non so]*

(c) Intervento divino *[Sì, no, non so]*

(d) Percezione extra sensoriale *[Sì, no, non so]*

(e) Intuizione (Presentimento) *[Sì, no, non so]*

(f) Un principio fisico non ancora scoperto dalla scienza *[Sì, no, non so]*

(g) Il fatto che tutto sia connesso nell'universo *[Sì, no, non so]*

Czech translation of **Questionnaire on Coincidence**

Bressan, P. (2002). The connection between random sequences, everyday coincidences, and belief in the paranormal. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 16(1), 17-34.

Jak často se Vám stává, že se v běžném životě, setkáváte se zvláštní nebo smysluplnou náhodou?

Prosím označte, jak často se Vám stávají následující situace na škále, kde 1= nikdy, 2=jednou či dvakrát, 3= několikrát 4= často, 5=velmi často

1) Souhra náhod může nastat v různých situacích. Jak často jste zažili takovou souhru náhod v každé z následujících kategorií?

(a) Po sobě jdoucí řady nebo shluky jmen, čísel či událostí stejného druhu (například to, když se setkáte opakovaně se slovem, které jste nikdy předtím neslyšeli, s odstupem několika málo hodin). *[1,2,3,4,5]*

(b) Spontánní asociace (například to, že na někoho či něco myslíte a brzy na tuto věc či člověka nečekaně narazíte). *[1,2,3,4,5]*

(c) Zážitky „malého světa“ (například to, že se setkáte s osobou, kterou člověk dlouho neviděl na nějakém velmi nepravděpodobném místě). *[1,2,3,4,5]*

(d) Vjemy něčeho vzdáleného ve vesmíru (například to, že se strachujete o někoho, komu se ve stejný čas, na jiném místě, stala nehoda). *[1,2,3,4,5]*

(e) Vjemy něčeho vzdáleného v čase (například to, že něco prožijete ve snu, a následně se to samé stane ve skutečnosti). *[1,2,3,4,5]*

(f) Neočekávané řešení problému (například to, že se setkáte s přítelem, který chce prodat svůj počítač přesně v okamžiku, kdy vy počítač sháníte). *[1,2,3,4,5]*

(g) Fenomén „Anděla strážného“ (například to, že zmeškáte pracovní pohovor a pak zjistíte, že to bylo lepší, protože dostanete mnohem lepší nabídku, kterou byste jinak zmeškali). *[1,2,3,4,5]*

2) Myslíte si, že takové souhry okolností / smysluplné náhody jsou dílem:

a) Úplné náhody *[Ano, Ne, Nevím]*

b) Osudu *[Ano, Ne, Nevím]*

c) Vyšší moci *[Ano, Ne, Nevím]*

d) Mimo-smyslového vnímání *[Ano, Ne, Nevím]*

e) Intuice *[Ano, Ne, Nevím]*

f) Fyzikálního principu dosud neobjeveného vědou *[Ano, Ne, Nevím]*

g) Skutečností, že vše souvisí se vším ostatním ve vesmíru *[Ano, Ne, Nevím]*

Italian translation of **Illusory Beliefs Inventory**

Kingdon, B. L., Egan, S. J., & Rees, C. S. (2012). The Illusory Beliefs Inventory: A new measure of magical thinking and its relationship with obsessive compulsive disorder. *Behavioural and Cognitive Psychotherapy* 40.1, 39–53.

Le voci sono valutate su una scala di 5 punti, da 1 (fortemente in disaccordo) a 5 (fortemente d'accordo).

1. Uso la preghiera per allontanare la sfortuna.
2. A volte ho cambiato i miei piani perché avevo un brutto presentimento.
3. L'anima non continua ad esistere dopo la morte
4. Credo nella magia.
5. A volte eseguo dei particolari rituali di protezione.
6. Se penso troppo a qualcosa di brutto, allora succederà.
7. Le forze magiche hanno avuto un impatto sulla mia vita.
8. E' solo questione di tempo prima che la scienza possa spiegare tutto
9. Faccio qualcosa di particolare per prevenire la sfortuna.
10. A volte ho la sensazione che possa succedere qualcosa, prima che questa accada.
11. Non è possibile lanciare un incantesimo.
12. La magia fa sì che i miracoli accadano.
13. La vita non è altro che una serie di avvenimenti casuali.
14. I portafortuna non funzionano.

15. Se penso troppo a qualcosa, allora accadrà.
16. Evito i numeri sfortunati.
17. La maggior parte delle cose che ci succedono sono il risultato del destino.
18. Credo che gli angeli custodi o altre entità spirituali mi proteggano.
19. La scienza è la chiave per comprendere come accadono le cose.
20. Anche solo i miei pensieri possono alterare la realtà.
21. C'è una forza invisibile che ci guida tutti.
22. Non si dovrebbe mai mettere alla prova il destino.
23. Non credo in una presenza spirituale.
24. Credo in una forza superiore o in Dio.

Czech translation of **Illusory Beliefs Inventory**

Kingdon, B. L., Egan, S. J., & Rees, C. S. (2012). The Illusory Beliefs Inventory: A new measure of magical thinking and its relationship with obsessive compulsive disorder. *Behavioural and Cognitive Psychotherapy* 40.1, 39–53.

from 1 (Naprosto nesouhlasím) to 5 (Naprosto souhlasím)

1. Používám modlitbu abych odehnal/a neštěstí.
2. Občas jsem už změnil/a své plány, protože jsem měl/a špatný pocit.
3. Po smrti duše nepokračuje v existenci.
4. Věřím v magii.
5. Občas provádím zvláštní rituály pro ochranu.
6. Když příliš přemýšlím o něčem špatném, stane se to.
7. Magické síly měly někdy dopad na můj život.
8. Je jen otázkou času, než věda dokáže vše vysvětlit.
9. Dělám něco zvláštního, abych předešel/předešla smůle.

10. Občas získám pocit, že se něco stane ještě předtím, než se to stane.
11. Není možné seslat kouzlo.
12. Kouzla způsobují, že se dějí zázraky.
13. Život není víc než řada náhodných událostí.
14. Talismany pro štěstí nefungují.
15. Když příliš přemýšlím o něčem špatném, stane se to.
16. Vyhýbám se nešťastným číslům
17. Většina věcí, které se nám dějí, jsou výsledkem osudu.
18. Věřím, že mě ochraňují strážní andělé nebo jiné duchovní síly
19. Věda je klíčem k porozumění, jak se věci dějí.
20. Samotné mé myšlenky mohou pozměnit skutečnost.
21. Všechny nás provází neviditelná síla.
22. Nikdy bys neměl/a pokoušet osud.
23. Nevěřím v duchovní přítomnost
24. Věřím ve vyšší sílu Boha.

Italian translation of **CREDs Exposure Scale**

Lanman, J. A., & Buhrmester, M. D. (2017). Religious actions speak louder than words: Exposure to credibility-enhancing displays predicts theism. *Religion, Brain & Behavior*, 7(1), 3-16.

Istruzioni: le seguenti domande riguardano le esperienze relative alla religione durante la tua crescita. Nello specifico, le domande riguardano le tue percezioni sui tuoi genitori o le principali figure di accudimento (come ad esempio tutori o parenti). Rispondi a ciascuna delle domande secondo la tua impressione complessiva sui tuoi genitori (o principali figure di accudimento), usando la seguente scala:

1 2 3 4 5 6 7 (from “in nessun modo” to “in misura estrema”)

1. In quale misura i tuoi genitori partecipavano a incontri e funzioni religiose?
2. In quale misura i tuoi genitori si impegnavano in attività di volontariato o beneficenza?
3. Complessivamente, in quale misura i tuoi genitori si comportavano come dei buoni modelli religiosi di riferimento?
4. Complessivamente, in quale misura i tuoi genitori facevano dei sacrifici personali per la religione?
5. In quale misura i tuoi genitori si comportavano onestamente con gli altri perché glielo insegnava la loro religione?
6. In quale misura i tuoi genitori vivevano una vita religiosamente pura?
7. In quale misura i tuoi genitori evitavano di far del male agli altri perché glielo insegnava la loro religione?

Czech translation of **Rational/Experiential Multimodal Inventory**

Norris, P., & Epstein, S. (2011). An experiential thinking style: Its facets and relations with objective and subjective criterion measures. *Journal of personality*, 79(5), 1043-1080.

from 1 (Naprosto nesouhlasím) to 5 (Naprosto souhlasím)

1. Užívám si problémy, které vyžadují usilovné přemýšlení.
2. Nejsem moc dobrý/dobrá v řešení problémů vyžadujících opatrnou logickou analýzu.
3. Užívám si intelektuální výzvy.
4. Preferuji komplexní problémy před jednoduchými.
5. Jsem nerad/a, když musím moc přemýšlet.
6. Řešení věcí opatrným uvažováním není mou silnou stránkou.
7. Nejsem příliš analytický myslitel
8. Snažím se vyhýbat situacím, které vyžadují o něčem přemýšlet do hloubky.
9. Přicházím na věci logicky mnohem lépe než většina lidí.

10. Mám logickou mysl.

11. Používání logiky obvykle dobře funguje při řešení problémů v mém životě.

12. Znát odpověď bez porozumění jejího důvodu mi stačí.

13. Užívám si četbu věcí, které vyvolávají vizuální obrazy

14. Užívám si představování si věcí

15. Umím si jasně představit nebo zapamatovat sochu nebo přírodní objekt (neživý), o kterém si myslím, že je velmi krásný

16. Silně se ztotožňuji s postavami ve filmech nebo v knihách, které čtu.

17. Mám sklony popisovat věci pomocí obrazů nebo metafor či kreativních přirovnání.

18. Umění je pro mě opravdu důležité.

19. Občas rád/a jen nečinně sedím a pozoruji dění

20. Mám oblíbené básně a obrazy, které pro mě hodně znamenají.

21. Když někam cestuji nebo řídím, vždycky pozoruji krajinu a scenérii.

22. Téměř nikdy nepřemýšlím v obrazech

23. Na mých emocích v mém životě příliš nezáleží.

24. Emoce opravdu moc neznamenají: přicházejí a odcházejí

25. Když mám silný emocionální zážitek, jeho efekt ve mně zůstane na dlouhou dobu.

26. Když jsem smutný/smutná, je to často velmi silný pocit.

27. Věci, které ve mě vyvolávají emoce, na ostatní zřejmě tolik nepůsobí.

28. Každodenní zážitky ve mně často vyvolávají silné pocity.

29. Raději bych byl/a občas naštvaný/naštvaná a občas šťastný/šťastná, než abych se neustále cítíl/a klidně.

30. Na děsivé filmy nebo knihy nereaguji tak emocionálně jako většina lidí.

31. Můj hněv je často velmi intenzivní.

32. Když jsem šťastný/šťastná, je to obvykle spíše pocit spokojenosti než rozjařenosti nebo vzrušení.

33. Rád/a se spoléhám na své intuitivní dojmy.

34. Často se řídím svými instinkty, když se rozhoduji o nějakém postupu.

35. Nemyslím si, že je dobrý nápad spoléhat se při důležitých rozhodnutích na intuici.
36. Věřím svým prvotním pocitům z lidí.
37. Mám sklony používat své srdce jako průvodce mých činů.
38. Užívám si učení skrze dělání něčeho, místo toho abych to nejdřív vymýšlel/a.
39. Často poznám, jak se lidé cítí, aniž by museli cokoliv říct.
40. Obecně se nespoléhám na své pocity, aby mi pomohly učinit rozhodnutí.
41. Popis zážitků opravdových lidí je pro mě přesvědčivější než diskuse o "faktech".
42. Nejsem moc spontánní člověk.

Czech translation of **Behavioral Inhibition System/Behavioral Activation System Scales**

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of personality and social psychology*, 67(2), 319.

1. Rodina je pro člověka nejdůležitější součástí života.
2. I když se mi má stát něco zlého, málo kdy zažívám strach nebo nervozitu.
3. Dělám, co můžu, abych dosáhl/a po čem toužím.
4. Když se mi v něčem daří, rád/a v tom pokračuji.
5. Jsem vždy ochotný/á vyzkoušet něco nového, když si myslím, že to bude zábava.
6. Je pro mne důležité, jak se oblékám.
7. Když získám něco, co jsem chtěl/a, cítím se vzrušený/á, nabuzený/á.
8. Když mě někdo zkritizuje nebo napomene, značně se mě to dotkne.
9. Když něco chci, obvykle do toho dám všechno.
10. Často dělám věci jen pro to, že by to mohla být zábava.
11. Je pro mě těžké najít si čas na věci, jako je návštěva holiče/kadeřnice.
12. Když vidím šanci získat něco, co chci, hned se do toho pustím.
13. Cítím se docela ustaraně nebo znepokojeně, když si myslím, nebo vím, že se na mě někdo zlobí.

14. Když vidím příležitost získat něco, co se mi líbí, hned se nadchnu.
15. Často dělám věci z popudu a bez přemýšlení.
16. Když si myslím, že se stane něco nepříjemného, obvykle jsem "jako na trní".
17. Často uvažuji nad tím, proč se lidé chovají tak, jak se chovají.
18. Když se mi stane něco dobrého, silně to prožívám.
19. Když si myslím, že se mi nepovedlo něco důležitého, cítím se ustaraně.
20. Toužím po vzrušení a nových prožitcích.
21. Když o něco usiluji, neohlížím se na pravidla, jdu do toho bez zábran.
22. Ve srovnání s mými přáteli mám jen velmi málo obav.
23. Výhra v soutěži by ve mně vyvolala vzrušení.
24. Mám obavy, že udělám něco špatně.

Czech translation of **Sexual Inhibition and Excitation Scales**

Carpenter, D., Janssen, E., Graham, C., Vorst, H., & Wicherts, J. (2011). Sexual Inhibition/sexual excitation scales-short form. *Handbook of sexuality-related measures*, 236-239.

1. Když se mě náhodně dotkne pro mě sexuálně přitažlivý neznámý člověk, snadno se sexuálně vzruším.
2. Pravděpodobně nedosáhnu silného vzrušení, když mám sex někde venku na skrytém místě a myslím si, že někdo může být nablízku.
3. Sexuálně se vzruším, když telefonuji s někým, kdo má sexy hlas.
4. Nemohu se sexuálně vzrušit, pokud se nesoustředím výhradně na sexuální dráždění.
5. Když jsem sám/sama a masturbuji, ale uvědomím si, že někdo může každou chvíli vejít do místnosti, mé sexuální vzrušení opadne.
6. Když si uvědomím riziko nákazy sexuálně přenosnou chorobou, je nepravděpodobné, že zůstanu sexuálně vzrušený/á.

7. Pokud by mě při sexuálních aktivitách mohl někdo vidět, je nepravděpodobné, že zůstanu sexuálně vzrušený/á.
8. Je pro mne snadné se vzrušit, když myslím na velmi atraktivní osobu.
9. Pokud dosáhnu sexuálního vzrušení (u mužů - mám erekci), chci co nejdříve pohlavní styk, než moje vzrušení opadne.
10. Když začnu mít sexuální představy, velmi snadno se vzruším.
11. Vidět jiné lidi při sexuálních aktivitách ve mně budí touhu se sexuálním aktivitám věnovat také.
12. Když mám rušivé myšlenky, mé sexuální vzrušení snadno opadne.
13. Když mě vyruší hudba, televize, nebo rozhovor, je nepravděpodobné, že budu i nadále sexuálně vzrušený/á.
14. Když se mnou flirtuje někdo přitažlivý, snadno se sexuálně vzruším.

Scientific Output

Chronological list of the scientific publications during my PhD study:

Binter, J., Boschetti, S., Hladký, T., Říha D. & Wells, T.J. (2020). Bolest a slast v sexuálním chování. *Sexuológia (Sexology)*, 1, 10 – 14.

Průšová, D., Boschetti, S., Hladký, T., Říha, D., Wells, T.J. & Binter, J. (2020). Vliv pornografie na přejímání emocí a sexuálního chování. *Sexuológia (Sexology)*, 2, 31 – 35.

Boschetti, S., Hladký, T., Říha D. & Binter, J. (2021). Sex a jídlo – slasti a cyklické stavy, které procházejí obdobnými procesy: jídlo kořeněné chilli a sex kořeněný sadomasochismem mají společné rysy. *Sexuológia (Sexology)*, 1, 31 – 38.

Prossinger, H., Boschetti, S., Hladký, T., Říha D. & Binter, J. (2021a). Konzumace pornografie obyvatel ČR a její vztah k demografickým proměnným: klastrování pomocí umělé inteligence. *Sexuológia (Sexology)*, 1, 7 - 16.

Prossinger, H., Hladky, T., Binter, J., Boschetti, S., & Riha, D. (2021b). Visual Analysis of Emotions Using AI Image-Processing Software: Possible Male/Female Differences between the Emotion Pairs "Neutral"–"Fear" and "Pleasure"–"Pain*"*. *In The 14th PErvasive Technologies Related to Assistive Environments Conference* (pp. 342-346).

Binter, J., Boschetti, S., Hladký, T., Prossinger, H., Wells, T. J., Jílková, J., & Říha, D. (2022a). Quantifying the Rating Performance of Ambiguous and Unambiguous Facial Expression Perceptions Under Conditions of Stress by Using Wearable Sensors. *In International Conference on Human-Computer Interaction* (pp. 519-529). Springer, Cham. doi: 10.1007/978-3-031-17615-9_36

Binter, J., Boschetti, S., Hladký, T., Prossinger, H., Wells, T. J., Jílková, J., & Říha, D. (2022b). A 'Serious Games' Approach to Decisions of Environmental Impact of Energy Transformation. *In International Conference on Human-Computer Interaction* (pp. 487-495). Springer, Cham. doi: 10.1007/978-3-031-05637-6_31

Boschetti, S., Prossinger, H., Hladký, T., Machová, K., & Binter, J. (2022). "Eye can't see the difference": Facial Expressions of Pain, Pleasure, and Fear Are Consistently Rated Due to Chance. *Human Ethology*, Volume 37, 046-072. doi: 10.22330/he/37/046-072

Kopecky, R., Příplatová, L., Boschetti, S., Talmont-Kaminski, K., & Flegr, J. (2022). Le Petit

Machiavellian Prince: Effects of latent toxoplasmosis on political beliefs and values. *Evolutionary Psychology*, 20(3). doi: 10.1177/147470492211126

Prossinger, H., Hladký, T., Boschetti, S., Říha, D., & Binter, J. (2022a). Determination of "Neutral"–"Pain","Neutral"- "Pleasure", and "Pleasure"–"Pain" Affective State Distances by Using AI Image Analysis of Facial Expressions. *Technologies,* 10(4), 75. doi: 10.3390/technologies10040075

Prossinger, H., Binter, J., Machova, K., Riha, D., Boschetti, S. (2022b). Machine Learning Detects Pairwise Associations between SOI and BIS/BAS Subscales, making Correlation Analyses Obsolete. *In Human Interaction & Emerging Technologies (IHIET-AI 2022): Artificial Intelligence & Future Applications*. vol 23. AHFE International, USA. doi: 10.54941/ahfe100903

Binter, J., Boschetti, S., Hladký, T. & Prossinger, H. (2023a, submitted) Ouch!" or "Aah!": Are Vocalizations of 'Laugh', 'Neutral', 'Fear', 'Pain' or 'Pleasure' Reliably Rated? *Human Ethology.*

Binter, J., Prossinger, H., Hladký, T., & Boschetti, S. (2023b, *submitted*). " Wow!" and" Aah": Bimodal Stimuli of Affective States with High and Low Intensity Displays. *Nature Scientific Reports*.

Binter, J., Hladký, T., Machová, K., Říha, D., & Boschetti, S. (2023c, *in press*). Validation of BIS/BAS, SIS/SES-SF and SOI-R Questionnaires in Two Czech Samples. *Czechoslovak Psychology*.

Boschetti, S., Binter, J. and Prossinger, H. (2023a, *in press*) Apoidolia: A New Psychological Phenomenon Detected by Pattern Creation with Image Processing Software together with Dirichlet Distributions and Confusion Matrices. *Proceedings in Adaptation, Learning and Optimization*.

Boschetti, S., Prossinger, H., Hladký, T., Říha, D., Příplatová, L., Kopecký, R. & Binter, J. (2023b, *in press*). Are Patterns Game for Our Brain? AI identifies individual Differences in Rationality and Intuition Characteristics of Respondents Attempting to Identify Random and Non-random Patterns. *In International Conference on Human-Computer Interaction*.

Boschetti, S., Prossinger, H., Prossinger-Beck, V., Hladký, T., Říha, D. & Binter, J. (2023c, *in press*). Never correct: The novel analysis of differing visual (facial expression) and acoustic (vocalization) bimodal displays of the affective states "pain", "pleasure", and "neutral". *In International Conference on Human-Computer Interaction*.

Kopecky, R., Boschetti, S., & Flegr, J. (2023, *submitted*). Effect of being religious on wellbeing in a predominantly atheist country: Explorative study on wellbeing, fitness, physical and mental health. *Scientific Reports*.

Prossinger, H., Boschetti, S., Říha, D., Pacovský, L. & Binter, J. (2023, *in press*) From Stone Age to New Age Statistics: How Neural Networks Overcome the Irreproducibility Problems in Choice Based Profile Creation. *In International Conference on Human-Computer Interaction.*

Collation of my publications used in my thesis

HCII INCONGRUENCE

Never correct: The novel analysis of differing visual (facial expression) and acoustic (vocalization) bimodal displays of the affective states “pain”, “pleasure”, and “neutral”

Silvia Boschetti [1[0000-0002-8048-4062]], Hermann Prossinger [1,2[0000-0002-1114-7687]], Violetta Prossinger-Beck[3[0000-0000-0000-0000]], Tomáš Hladký[1[0000-0001-5106-5902]], Daniel Říha[1[0000-0001-5142-4485]], and Jakub Binter[1[0000-0001-5304-2130]]

[1] Faculty of Humanities, Charles University, Prague, Czech Republic
[2] Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria
[3] Dresden International, TU Dresden, Dresden, Germany
silvia.boschetti@natur.cuni.cz

Abstract. In real-world scenarios, humans estimate a large proportion of their perceived world contextually and use previous information to adjust or modify their expectations and responses. A typical example of congruence is when a smile (visual modality) is accompanied by a laugh (acoustic modality). Pain and pleasure are extremely intensive affects, rarely correctly assessed. It rarely happens that the affective communication is incongruent in different modalities. Intentional combinations of these two affective expressions may be implemented by using audiovisual media. It is to be expected that presentations involving incongruent combinations during gaming and other interactions with machines will alter perceptions and impact of ambiguity. To evaluate the impact of sensory crosstalk a novel statistical analysis was developed to estimate impact of each modality. The results show that for the pairing pain/pleasure, the visual modality dominates. The acoustic modality dominates for pain/pleasure, pain/neutral, and pleasure/neutral. For neutral/pain and neutral/pleasure neither the visual nor the acoustic modality enables highly correct ratings. The findings are of high value to psychology of perception on theoretical level and game/media developers as application fields.

Keywords: Visual Displays, Acoustic Displays, Affective states, Dirichlet Distribution, Confusion Matrix, Heat Maps.

1 Introduction

Consider an observer watching a single frame extracted from a video and hearing the simultaneous audio signal. In such a case, the person is observing (seeing and hearing) a bimodal signal.

If the observer is asked to rate the observation, and the researcher is interested in whether the observer relies on the visual or the acoustic modality when rating, the researcher can modify the presentation. The observer is presented with a video frame as a visual modality together with an audio signal as the acoustic modality, but these two presented modalities are different. The observer has to rely on one of the two modalities. As the researcher knows the visual modality and the acoustic one, he/she can conclude on which modality the observer relied when rating — never on both, because the presented modalities exclude each other. We call such a set-up an incongruent bimodal presentation.

In this paper, we present many women as well as many men (separately) with incongruent bimodal stimuli, namely the pairings of the affective states pain, pleasure and neutral (speech) in all six combinations. These were presented (separately) by males and females, which we call stimuli. Details are described in the Methods section. Nomenclature: we henceforth call the participating observers raters, and the presenting females and the males from which the presentations of the bimodal affective states were derived as female or male stimuli.

In the set-up presented here, the visual and auditory information does not aggregate correctly. One of the few comparable real-life experiences we could find, which gave the name to the perceived phenomenon, is ventriloquism. This ancient art of making it seem as if the voice originates from a different source than the (visually presented) mouth can be traced back to Greek and Roman antiquity

and is still used in puppet theaters [Dumbstruck, 2000].
Since the invention of recording devices for visual displays (recordings by video cameras, for instance) and those for acoustic accompaniment (microphones attached to recording devices), the possible experimental situation has dramatically changed. We can separate the two recordings and match them incongruently.
There are three incompatible hypotheses which we can test by comparing incongruent recordings: (1) The more intensive affective display (vocalization or facial expression) dominates over the less intensive one. (2) The unambiguous input (irrespective of the intensity) will dominate over the ambiguous one. (3) The hypothesis predicted by the sensory dominance theory [Hutmacher, 2019], which expects one sensory input to always be dominant over the other one, irrespective of the displayed intensity or the modality.

2 Materials and Methods

Incongruent pairings (such as the visual stimulus pain with the acoustic stimulus neutral speech) were presented to 902 raters (526 women and 376 men, aged 18–50 years). The raters were to rate whether an incongruent pairing was perceived as positive (for pleasure), negative (for pain) or neutral (for neutral speech) using keyboard or touch-screen. The data were collected in the Czech Republic in 2021 via the agency Czech National Panel (narodnipanel.cz) and a science-oriented online portal pokusnikralici.cz using the online platform for data collection Qualtrics™. Prior to their commencing to rate, the participants supplied their (biological) sex and their age to the nearest year. Criteria for inclusion were: (a) age of raters between 18 and 50 years, and (b) at least a minimal experience with adult media, since the displays and the vocalizations used in this study were extracted from such materials.
From the numerous audio-visual materials, ten were chosen (five with female expressers and five with male expressers). Based on the plot in each of these ten, three frames with faces and three audio tracks were selected (one each of neutrality, pleasure, and pain). The two highly intensive facial expressions (pain and pleasure) were also tested for difference by using an Artificial Neural Network algorithm [Prossinger et al., 2021].
We used KDE (Kernel Density Estimation) to find the distribution of ages of the female and the male raters, separately. We then determined the estimators, in particular the $HDI_{95\%}$ (the Highest Density Interval at 95% significance [Kruschke, 2015]). A large overlap of this interval documents no significant difference in the age distributions.
We tallied whether the rating was correct because the visual modality was rated correctly, or whether the acoustic modality was rated correctly or whether neither modality was rated correctly (which we tallied as ‘both incorrect’).
For every rater sex and for every stimulus sex and each bimodality, the chosen responses are Dirichlet distributed. There are $2 \times 2 \times 6 = 24$ such distributions. For each, we calculated the mode (a 3D vector). The entries in the columns (i.e. the suite of the six pairings for female raters of the acoustic stimuli, say) male versus female are also Dirichlet distributions. We used Monte Carlo methods for pairwise comparisons.
Lastly, we determined whether female raters of stimulus for one sex for a specific modulus significantly differs from that of another sex. These comparisons are Beta distributions; the detailed description of the method of finding the significance is in the Appendix. We repeated these significance calculations for the other rater sex and then also for the mixing of the sexes of the raters with the sexes of the stimuli.

Table 1. The descriptors and the estimators for the ages of the raters, separated by (biological) sex. $\mathbb{E}$ is the expected value (estimated by $\mathbb{E} = \int_0^{60} u \times pdf(\text{KDE}, u)\, du$, where KDE is the kernel density estimation), and $HDI_{95\%}$ (highest density interval: the interval with a 95% probability of observing an age; [Kruschke, 2015]). We observe that for both the male and the female raters, the mode is considerably different from the expected age $\mathbb{E}$. The *pdf*s of the KDE distributions and the histograms are in Fig. 1.

Estimator	Male Raters	Female Raters
N	376	526
Range (years)	18–50	18–50

Mode (years)	41.7	27.0
$\mathbb{E}$ (years)	33.6	30.9
Mean (years)	33.6	30.9
$HDI_{95\%}$ (years)	16.4–50.3	15.8–49.8

3 Results

Fig. 1 and Table 1 show that the distributions of the male and female raters' ages have different modes. We use KDE (kernel density estimation) with a Gaussian kernel because, as Fig. 1 shows, it is not to be expected that raters of either sex have a parametric distribution or even a superposition of one or two such parametric distributions. We also note that modes and expectation values differ between the rater sexes and, furthermore, modes and expectation values differ for the raters of the same sex. The $HDI_{95\%}$ uncertainty interval is very broad, so we can consider the distributions for both rater sexes to be comparable to a uniform distribution of respective ages. The confusion matrix shows that the two distributions are not significantly different.

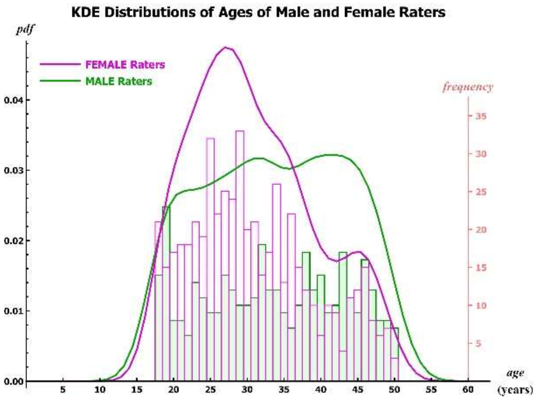


Fig. 1. The distributions of the ages of males and of females estimated using a KDE (kernel density estimation) with a Gaussian kernel. Superimposed of the graphs of the two *pdf*s are the two histograms of the registered ages, separated by sex (scale of frequencies on the right). Modes, expected values, means, and $HDI_{95\%}$ of both distributions are listed in Table 1. The two distributions are not significantly different; the confusion matrix is $\begin{pmatrix}46.3 & 53.7\\ 28.3 & 71.7\end{pmatrix}\%$ (the method of computing this confusion matrix is described in [Boschetti et al., 2022]).

For the column comparisons for each modality (Fig. 2), no significant difference was observed. Fig. 3 shows two examples: the two Dirichlet distributions for two sets of ratings. The extent of overlap between the *pdf*s of these indicates the significance level. In Fig. 3, the overlap is extraordinarily small, so the differences in rating (the off-diagonal entries in the confusion matrix) are highly significant. All the off-diagonal entries in all the confusion matrices were much less than 1×10^{-3} (not shown), therefore, for each comparison for each modality there is a significant different in rating distributions.

For the visual modality, we observe that one bimodal affect display (pleasure with pain) is markedly different from all other affective displays (Fig. 2). For the acoustic modality, we observe that the pairings pain with pleasure, pain with neutral and pleasure with neutral are significantly different from all other pairings. For the incorrect ratings (if raters chose neutral, say, when the visual display is positive and the acoustic display is negative) only the pairing neutral with pleasure is significant.

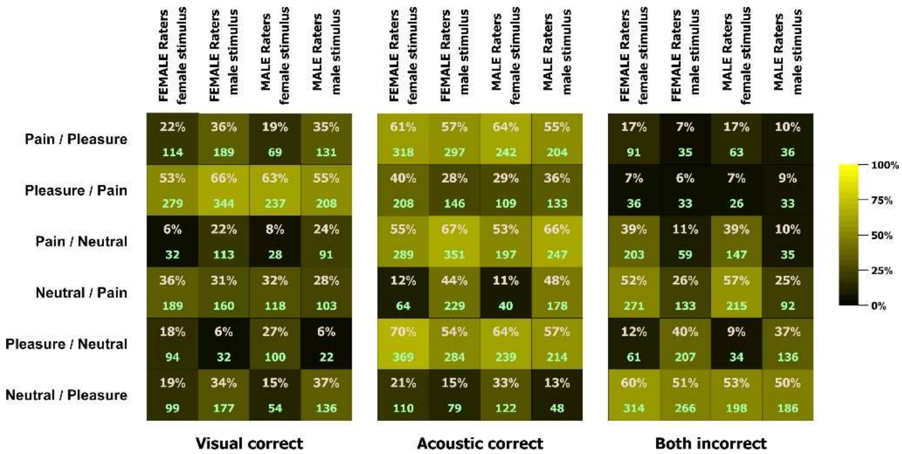


Fig. 2. The heat map of the distributions of incorrect (or possibly guessed) ratings by male and female raters of male and female stimuli. The numbers (light cyan) in the squares are the number of occurrences. Color-coding of the background is of the fraction (in percent) of the occurrences, also displayed as off-white numerical values. We observe that for the pairing pain/pleasure the visual modality dominates. We also observe that the acoustic modality dominates for pain/pleasure, pain/neutral, and pleasure/neutral. For neutral/pain and neutral/pleasure, on the other hand, neither the visual nor the acoustic modality enables highly correct ratings.

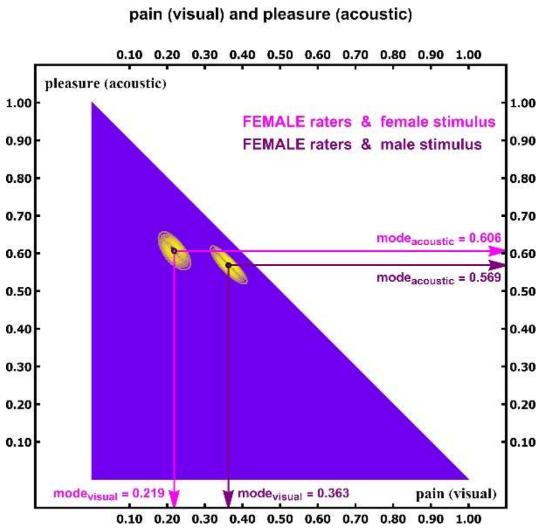


Fig. 3. The likelihood functions of the Dirichlet distributions of the bimodal pairings of pain/pleasure of the female and male stimuli, as rated by the female raters. The probabilities are s_1 (pain) and s_3 (pleasure). The likelihood function is defined over the triangle shown in purple, because $s_1 + s_2 + s_3 = 1$. Modes are shown; these are the entries in Fig. 2. Contours (yellow, are in steps of $\frac{1}{14}\mathcal{L}_{\max}$. The contours do not overlap, clearly showing that the rating distributions are significantly different.

We conclude that for the pleasure/pain pairing, the visual modality is far superior to the acoustic modality (namely, that of pain). For the three pairings pain/pleasure, pain/neutral, and pleasure/neutral, the acoustic modality dominates. To summarize: Fig. 2 shows that the visual modality pleasure dominates when paired with pain as an acoustic modality, whereas for the acoustic modality, pleasure dominates with the visual modality pleasure, and the acoustic modality neutral dominates when paired with pleasure and with pain. Only for the pairing neutral pleasure were the raters (of either sex) incapable of rating the dominant modularity correctly (Fig. 2).

For the visual modularity, the comparison of the female ratings of male versus female stimuli was significantly different from than that of the male raters rating these two stimuli (Fig. 4). This significant difference (again for the visual modularity) was also observed for pain/neutral, pleasure/neutral and neutral/pleasure.

For the acoustic modularity, the comparisons of female versus male raters showed a significant difference for pain/neutral, neutral/pain and neutral/pleasure (Fig. 4). Visually, no combinations of raters and stimuli showed a significant difference.

We observe no pattern of significant differences for male versus female raters for a given stimulus (Fig. 4); neither visually nor acoustically. However, there are three exceptions; female versus male raters rate the female stimuli significantly differently for one visual pairing (pleasure/neutral) and two acoustic pairings (pleasure/pain and pleasure/neutral).

4 Discussion

Because of the presented incongruence, no rater could rate both presented modalities correctly, but could rate both incorrectly (because there existed a third choice/option).

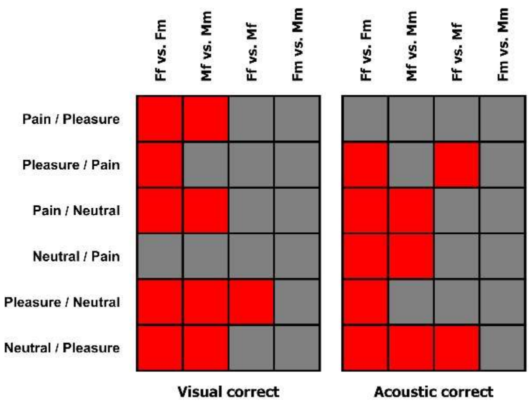


Fig. 4. A heat map of the distributions of significant differences between female (**F**) versus male (**M**) raters rating female (**f**) or male (**m**) stimuli in all eight combinations. The bright red squares display significant differences. Significance has been determined by calculating whether the mode of the Beta distribution of the comparison is significantly different from $s = \frac{1}{2}$, based on the entries displayed in Fig. 2. (The method is explained in the Appendix). For $\frac{20}{48} = \frac{5}{12} = 41.6\%$ of the combinations there are significant differences. We observe that for neutral/pain the acoustic modality leads to a significantly different rating by female raters rating female stimuli versus male stimuli and, likewise, by male raters rating female stimuli versus male stimuli. For pleasure/pain, the female raters' ratings of female stimuli versus male stimuli are significantly different for both the visual and acoustic mode. For male raters, this is not the case.

The patterns we found document that hypothesis (1) is not rejected, but (2) and (3) are rejected. The study of the ability to correctly attribute emotional and affective expressions of others was impacted by recent papers showing counterintuitive evidence that affective states with high intensity are harder to correctly assess from facial expression [Aviezer et al., 2012; Boschetti et al., 2022; Binter et a., 2023]. This is confirmed (Fig. 2); the acoustic modality allows for reliable ratings. When extremely intensive facial and vocal displays are rated, the accuracy is usually low, since the stimuli appear inherently ambiguous [Aviezer et al., 2012; Wenzler et al., 2016; Boschetti et al., 2022; Binter et al., 2023]. We find that whenever there is a contradictory bimodality involved, then the ambiguity is suppressed — a novel, unexpected finding.

Only in one pairing of incongruent affective states with incongruent modularity do we find evidence of sensory crosstalk: when the visual modularity makes the rating of acoustic modularity impossible. The raters are confused for the pairing neutral (visual) with pleasure (acoustic).

Although raters of either sex are consistent in ratings based on the visual modularity (Fig. 2), they rate the affective states displayed by stimuli of different sex significantly differently in $\frac{2}{3}$ of the pairings. For the acoustic modularity, in $\frac{1}{2}$ of the pairings, there are significantly different ratings of the displayed incongruent modularity.

5 Conclusion

The alteration of perception can be used for intentional modifications of scenarios as a 'twist' (i.e. a novel ambiguity) which can be, if well placed, cause unexpected effects on players (consumers and users) of the gaming product. On the other hand, we know from previous research, that pain and pleasure were overwhelmingly incorrectly attributed in congruent pairings. The impact of intentional incongruence generating ambiguity may be specifically implemented by the creators of gaming scenarios (for example, characterizing versus mischaracterizing a villain in crime-type games). Our investigations contribute to communicating to creators of gaming scenarios the effectiveness — or ineffectiveness — of incongruent pairings of two affects and a natural expression. Furthermore, there is an implication for the field of perception psychology since this unexpected result can be only found using the novel methodological approach developed by one of the authors (HP) for this very reason.

Ethics Statement

The project was evaluated and approved by the Ethical Committee of the Faculty of Science, Charles University, as part of broader project (7/2018). GDPR regulations were followed at all times.

Funding and Acknowledgement

The research and data collection was supported by Czech Science Foundation (Grant No. 19-12885Y "Behavioral and Psycho-Physiological Response on Ambivalent Visual and Auditory Stimuli Presentation").

Conflicts of Interest

The authors declare no conflicts of interest.

References

Bertelson, P, Aschersleben, G. (2003): Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *Int J Psychophysiol*. 50 (1–2): 147–155. doi: 10.1016/S0167–8760(03)00130–2.

Boschetti, S., Prossinger, H., Hladký, T., Machová, K., Binter, J. (2022) "Eye can't see the difference": Facial Expressions of Pain, Pleasure, and Fear Are Consistently Rated Due to Chance. *Human Ethology*, 37, 46–72. Doi: 10.22330/he/37/046-072.

Binter, J., Boschetti, S., Hladký, T., & Prossinger, H. (2003) "Ouch!" or "Aah!": Are Vocalizations of 'Laugh', 'Neutral', 'Fear', 'Pain' or 'Pleasure' Reliably Rated? Preprint at https://psyarxiv.com/rf7vw/.

Brandner, J. L., Pohlman, J., & Brase, G. L. (2021). On hits and being hit on: Error management theory, signal detection theory, and the male sexual overperception bias. *Evolution and Human Behavior*, *42*(4), 331–342.

Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, 81(3), 429–450. doi: 10.1007/s10472-017-9564-8.

Connor, S. (2000) Dumbstruck: A Cultural History of Ventriloquism, Oxford University Press, Oxford, UK.

de Gelder B, Pourtois G, Weiskrantz L: (2002) Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proc Natl Acad Sci USA*. 99 (6), 4121–4126. doi: 10.1073/pnas.062018499.

Gregorić, B., Barbir, L., Ćelić, A., Ilakovac, V., Hercigonja-Szekeres, M., Perković Kovačević, M., ... & Heffer, M. (2014). Recognition of facial expressions in men and women. *Medicina Fluminensis: 50*(4), 454–461.

Hess, U., Blairy, S., Kleck, R.E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior*, *21*, 241–257.

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? Frontiers in psychology, 10, 2246.

Kruschke, J.K. (2015) Doing Bayesian Data Analysis. A Tutorial with R, JAGS, and STAN. 2nd edn. Elsevier/Wiley, Waltham, MA, USA.

Perilloux, C., Easton, J.A., Buss, D.M. (2012). The misperception of sexual interest. *Psychological Science*, *23*(2), 146–151.

Pourtois G, de Gelder B, Vroomen J, Rossion B, Crommelinck M. (2000) The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, 11 (6): 1329–1333. doi: 10.1097/00001756–200004270-00036.

Pourtois G, de Gelder B, Bol A, Crommelinck M. (2005) Perception of facial expressions and voices and of their combination in the human brain. *Cortex* 41(1): 49–59. doi: 10.1016/S0010-9452(08)70177-1

Shimojo S., Shams L. (2001) Sensory modalities are not separate modalities: plasticity and interactions. *Curr Opin Neurobiol* 11(4) 505–509. doi: 10.1016/S0959-4388(00)00241-5

Vroomen J., Driver J., de Gelder B. (2001) Is cross-modal integration of emotional expressions independent of attentional resources? *Cogn Affect Behav Neuroscience* 1(4) 382–387. doi: 10.3758/CABN.1.4.382

Wada Y., Kitagawa N., Noguchi K. (2003) Audio-visual integration in temporal perception. *Int J*

Psychophysiol 50(1–2) 117–124. doi: 10.1016/S0167-8760(03)00128-4.

Appendix

We show a method of determining whether two sets of ratings are significantly different with an example (Fig. A-1). The rating entries of the female raters for the male stimuli are n_1 and the entries for the male raters for the male stimuli are n_2; then the Beta distribution is $\mathcal{Be}(n_1+1, n_2+1)$. The two boundaries $\left[\frac{1}{2}, u_{\text{upper}}\right]$ of the HDI (Highest Density Interval) are determined by $pdf\left(\frac{1}{2}\right) = pdf(u_{\text{upper}})$. (Comment: solving for u_{upper} requires computing power.) The probability of HDI is determined by

$$\text{HDI}_{\text{probability}} = \int_{\frac{1}{2}}^{u_{\text{upper}}} pdf(s)ds$$

Comment: this integral can be easily computed using the CDF (Cumulative distribution function) of the Beta distribution: $\text{probability} = \text{CDF}\big(\mathcal{Be}(n_1+1, n_2+1), u_{\text{upper}}\big) - \text{CDF}(\mathcal{Be}(n_1+1, n_2+1), \frac{1}{2})$. If the computed probability is less than 95%, then the significance level is greater than 5%. In this case, the deviation of the mode from $\frac{1}{2}$ is insignificant and the differences between the rating distributions are insignificant. In the example, shown in Fig. A-1, the probability is 60%; the significance level is therefore 40% and the observed difference in the rating distributions is insignificant.

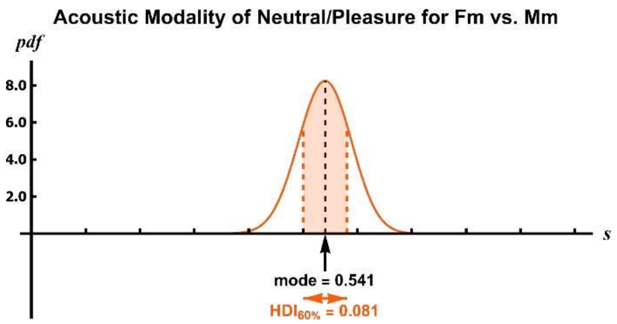


Figure A-1 A graph showing the method of determining the significance of the difference between two ratings of the same modality by raters of different sex (female **F** versus male **M**) of the male stimuli **m**. In this example, the analysis for the acoustic modality is shown, and the rating numbers of female and male stimuli are the concentration parameters of the Beta distribution (each increased by +1). s is the (Bayesian) probability. The mode of the distribution is shown, flanked by the upper and lower bounds of $\text{HDI}_{60\%}$. The dashed orange lines indicate the equal likelihoods. The area shaded in orange is the probability that the mode is significant, in this case $100\% - 60\% = 40\%$. The mode is not significantly different from $\frac{1}{2}$, and the ratings by the females and the males of the male stimuli is not significantly different at the 5% significance level.

Are Patterns Game for Our Brain? AI identifies individual Differences in Rationality and Intuition Characteristics of Respondents Attempting to Identify Random and Non-random Patterns

Silvia Boschetti[1[0000-0002-8048-4062]], Hermann Prossinger[1,2[0000-0002-1114-7687]], Tomáš Hladký[1[0000-0001-5106-5902]], Daniel Říha[3[0000-0001-5142-4485]], Lenka Příplatová[1[0000-0002-2860-3853]], Robin Kopecký[1[0000-0002-7140-4649]], and Jakub Binter[1[0000-0001-5304-2130]]

[1] Faculty of Sciences, Charles University, Prague, Czech Republic
[2] Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria
[3] Faculty of Humanities, Charles University, Prague, Czech Republic
silvia.boschetti@natur.cuni.cz

Abstract. In our everyday life we rely on set of heuristics that involve estimation of meaningful connections between events. In human evolutionary history, it was less costly to overestimate the meaning. The psychological phenomenon of apophenia (overperception of patterns) is then an adaptive response. It may manifest also as overperception of visual patterns (pareidolia). The underperception was rarely studied and researchers mainly used unsuitable stimuli sets for the purpose. After researching this phenomenon using patterns with transparency, geometric shapes, and color (Boschetti et al., 2023), we developed new set of black and white high-contrast stimuli. These were presented to participants four times in different orientations to limit guessing to 6% chance. Using ANN (Artificial Neural Networks), we associated the responses to the Rational Experiential Multimodal Inventory Subscales Rationality and Intuition. We were able to identify two clusters for each subscale and found associations of the participant responses with pattern identification success (or lack thereof). Our discoveries extend previous findings concerning this phenomenon and provides us with a foundation for constructing and designing artificial environments with high attention to cues given to users.

Keywords: Random Patterns, Non-random Patterns, Rationality, Intuition, Clustering Algorithms, Artificial Neural Networks, Apophenia, Apoidolia.

6 Introduction

The ability to correctly detect and recognize patterns is fundamental for human while they interact with the environment irrespective of whether it is real or virtual. The ability to distinguish between what is relevant and meaningful and what is not is an essential part of the decision-making process that underlies every behavior: the more accurate the detection of an existing pattern the more effective the decision among behavioral options.

However, in terms of cost-benefit, false positive (detection of a pattern in a random distribution) and false negative (failure in detecting a pattern when one is actually present) errors are very different: Error Management Theory (EMT) claims that over-detection of patterns is safer than under-detection (Johnson, 2009). For this reason, the bias for over-detection of pattern (called apophenia) is considered an advantage and is, arguably, an evolved mechanism in humans (Barrett, 2000).

The cognitive and perceptual bias is a tendency that construes how we perceive and interpret the world. The subjective sensitivity to pattern detection can lead to the miss-attribution of agency to non-living objects (Hyperactive Agency Detection Device) and ultimately to supernatural beliefs (Barrett, 2000). Indeed, cognitive and perceptual biases are candidate for explaining the emergence and the perseverance of paranormal beliefs (Willard & Norenzayan, 2013). Previous studies have associated the overperception of patterns with different types of beliefs in the supernatural, beliefs in the paranormal and assigning meaning to coincidence of events that are only due to chance (Zhou & Meng, 2020; Bressan, 2002).

The specific type of apophenia is the perception of a pattern in visual stimuli and is called pareidolia. Most of the work on pareidolia involves the overperception of faces, which are a very specific type of visual stimulus (Tsao & Livingstone, 2008) and the results of such studies may not be generalizable to stimuli different from facial ones. Another problem occurs when the stimuli chosen are naturally occurring (i.e. clouds, stains). Is it possible to control the extent of naturalness present or absent in an imaged pattern? Furthermore, even though such stimuli are suitable for detecting pareidolia (as false positive error in perception), they cannot be used to detect the false negative error (called apoidolia; Boschetti at al., 2023). Both apoidolia and pareidolia have been found to be more strongly related to thinking styles than to religious beliefs or belief in magic. This outcome was found in a previous study that used colored stimuli created by a multidimensional random number generator to produce maps of random colored squares underlined with geometrical figures as patterns (Boschetti et al., 2023). Specifically, the REIm (Rational/Experiential Multimodal Inventory; Norris & Epstein, 2011) was designed to measure the functionality of the experiential system of reasoning (based on the associative processes heuristic and intuitive) and the analytical system (based on high cognitive and meta-cognitive abilities and rule-based processing).

For this study we used two subscales from the REIm: one measuring *rationality* and the other measuring *intuition*. Furthermore, instead of colored stimuli we used black and white stimuli reminiscent of QR codes (Fig. 1), thereby maximizing the visual contrast in the elements of the stimulus. There is one further (important) limitation of the previous studies, namely, *The Lady Tasting Tea* problem (Fisher, 1956): how to avoid the observer from guessing the correct answer; more specifically, minimizing the probability of supplying a correct answer merely due to chance.

This simple, straightforward perceptual test presented here can be used as a diagnostic tool for users and profile creators when intending to maximize the experience in human-computer interactions. It also allows for addressing complex psychological phenomenon while limiting the number of response options.

7 Materials and Methods

7.1 Participants

A total of 174 participants from Italy and the Czech Republic responded to the queries about Rationality and Intuition. They (henceforth called respondents) also attempted to identify which of the eight presented distributions of black and white squares (henceforth called patterns) were random and which were non-random (henceforth called random and non-random patterns).

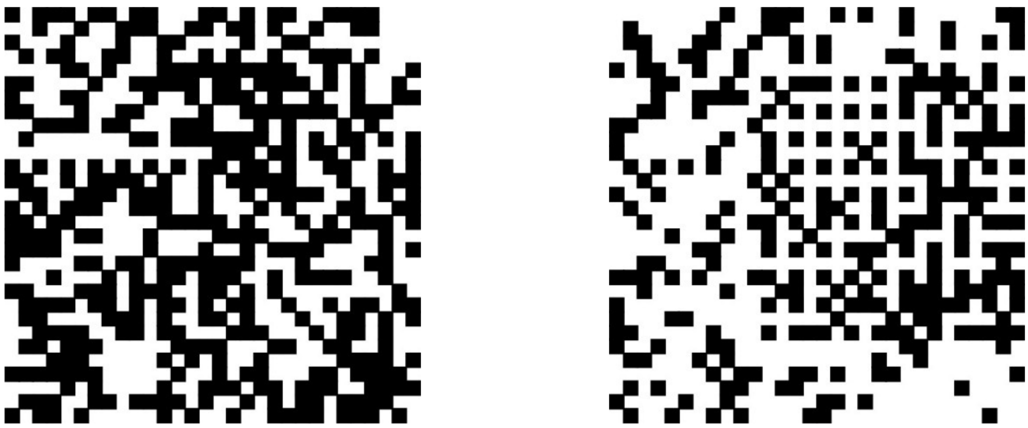


Fig. 2. Two of the eight patterns presented to the respondents. For every rectangle presented, they were asked to identify which pattern is random and which is nonrandom. The random patterns were presented four times, once as generated by the (quasi) random number generator, once after mirroring across the horizontal axis, once after mirroring across the vertical axis and once after rotating by 90°. Likewise, for the nonrandom pattern. The eight patterns were presented in randomized order, so as to prevent any possible memory effects. The reason for presenting each random and each non-random pattern four times is to suppress *The Lady Tasting Tea* effect (details in the text).

7.2 Stimuli

Four random patterns were generated, as well as four non-random ones (two are shown in Fig. 1). Respondents were asked to identify which of the patterns were random and which were not. We presented each pattern four times in order to avoid *The Lady Tasting Tea* problem (if a two-option challenge is presented to a respondent, the chance of correct identification is $\frac{1}{2}$; this chance is far too high; if the challenge is presented four times in random order, then chance of correct identification without guessing is $\frac{1}{2^4} = \frac{1}{16} \sim 6\%$; Fisher, 1956). We did not repeat the presentations 5 times, because of possible uncontrollable effects due to the tiring of the participants.

7.3 Questionnaire

We used the Rational/Experiential Multimodal Inventory (REIm; Norris & Epstein, 2011) to measure two aspects of thinking styles, namely analytic/rational and intuitive (specifically an automatic/experiential approach to construction and interpretation of the perceived world). The questionnaire is composed of four independent subscales, one for a rational system and three for different aspects of the experiential system. The subscales used in the present study are: Rationality (12 queries) and Intuition (10 queries); both are listed in the Appendix. We narrowed down our analysis to these two subscales because, from a theoretical prospective and from the results of a previous study (Boschetti et al., 2023), they are the most relevant. The responses were on a 5-point scale ranging from R1 (Completely False) to R5 (Completely True); we note that none are cardinal numbers.

7.4 Statistical Methods

The responses to the queries are ordinal numbers, not cardinal numbers (Blalock, 1960). They may not be directly converted into cardinal numbers, because a change in the choice of the mapping results in a change the statistical signal. (The statistical analysis is then the statistical analysis of the mapping, not of the data). Rather, the responses must be mapped into unit vectors. If a response to a query is 'B', say, then the vector is $(0 \quad 1 \quad 0 \quad 0 \quad 0)^T$. If a respondent identifies a pattern as random, the response is $(1 \quad 0)^T$, and if it is non-random, then the response is $(0 \quad 1)^T$. The response vectors are then concatenated; the resulting vector is called a feature vector (Murphy, 2012). For Rationality and pattern identification, the dimension of the feature vector is $4 \times 2 + 4 \times 2 + 12 \times 5 = 76$. This feature vector has interdependencies (perhaps nonlinear ones). In order to avoid the 'curse of dimensionality' (Bellman, 1961), dimension reduction is necessary. We used an artificial neural network (specifically, an autoencoder) to dimension reduce the feature vector. The autoencoder will also detect nonlinear interdependencies among the components of the feature vectors. We then use a clustering algorithm (in our case, DBSCAN; Ester et al., 1990) to identify the clusters in the space of the dimension-reduced feature vectors.
We can 'backtrack' the mapping and identify which points in the dimension-reduced vector space 'belong' to the respondents' feature vectors. We use these identifications for further analyses. AC (Correspondence Analysis; Benzécri, 1973; Greenacre, 2007; Beh & Lombardo, 2014) of the contingency table (pattern identification versus thinking style) allows us to determine whether there exist associations and, if so, how many.
We repeat the above procedure for the combination of responses to the queries of the Intuition questionnaire and the *same* identification of random and non-random patterns by the participants.

8 Results

8.1 Pattern Identification

We observe that only 3 (1,7%) of the 174 participants succeeded in correctly identifying all 4 nonrandom and all 4 random patterns. Some respondents succeeded in correctly identifying all random patterns, but they did not correctly identify all non-random patterns. Others succeeded in correctly identifying all non-random patterns, but did not succeed in identifying all random patterns. Many respondents succeeded only partially. Most respondents (140; 65% or close to $\frac{2}{3}$) did not succeed in correctly identifying any non-random patterns and any random patterns. We discovered that the relations between the extent of correctly identifying patterns are not independent of the responses to queries; this is valid for both Rationality and Intuition. In fact, cluster analysis shows that in both cases of thinking style (Rationality and Intuition), there are two distinct clusters (Fig. 2).

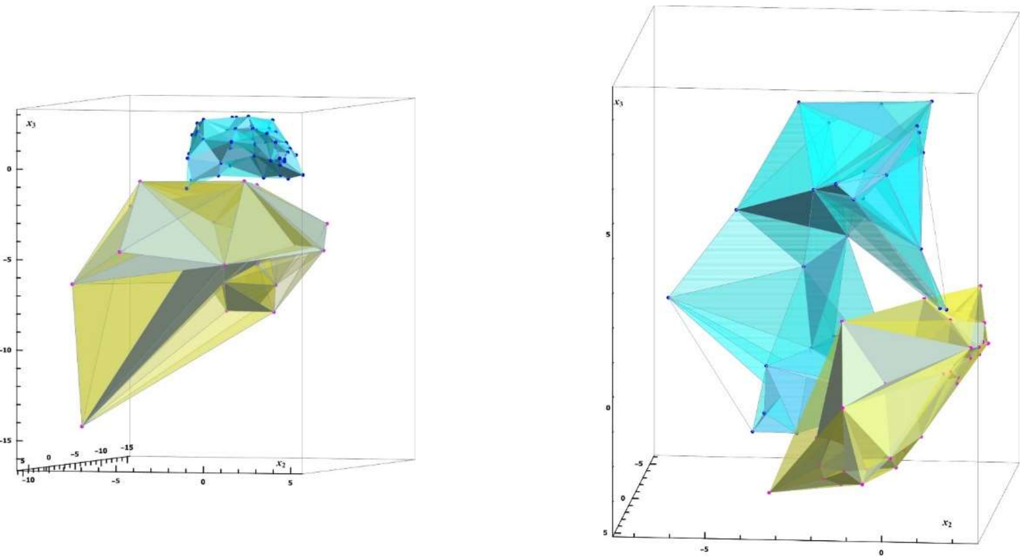


Fig. 3. The two clusters obtained after dimension-reducing the feature vectors using an autoencoder for Rationality (left) and Intuition (right). In both, the points are the coordinates of the dimension-reduced feature vector of each participant's responses to the Rationality or Intuition queries together with the identification of a random or nonrandom distribution of squares. The concave hull connects the points in each of the clusters. For Rationality, Cluster #2 has 39 points and Cluster #1 has 174 points and for Intuition, Cluster #1 has 38 points and Cluster #2 has 175 points. The dimension reduction to 2D did not separate clusters, but dimension reduction to 3D did. The axes are numerical only; they are not directly interpretable.

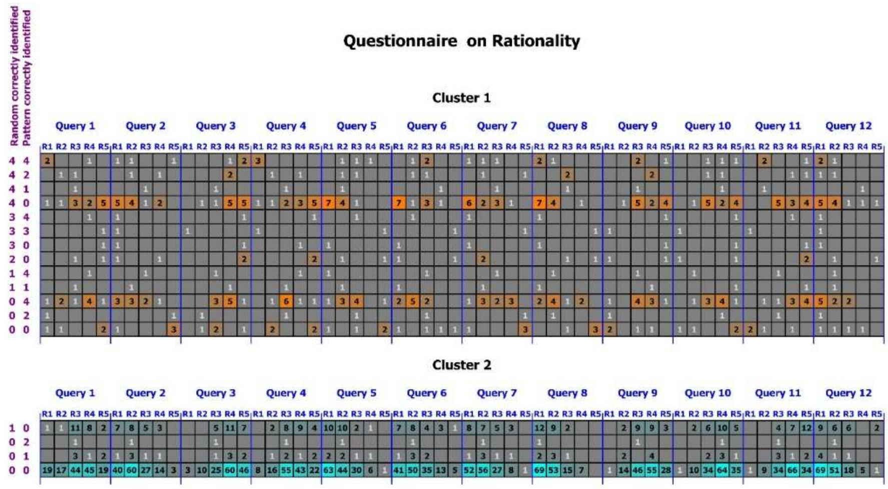


Fig. 4. The distribution of the nonzero entries of nonrandom and random pattern identification versus responses to the rationality queries. Those rows that have no nonzero entries have been deleted. The brightness of the colors identifies the percentages of occurring entries for a response (this fraction is normalized according to the total number of respondents in a cluster). In Cluster #2, for instance, none of the respondents correctly identified more than one random pattern and the greatest majority did not succeed in distinguishing between random and nonrandom patterns. In Cluster #1, on the other hand, many respondents were able to identify random patterns and many (other) respondents could identify nonrandom patterns, without successfully identifying nonrandom patterns. The few respondents who did not succeed in identifying either random or nonrandom patterns (who are expected to be in Cluster #2) are part of this cluster because the autoencoder also dimension reduces the feature vector, which contains the query responses. There are 3 respondents (1.7% of all 174) who correctly identified both the random patterns and the nonrandom patterns.

8.2 Associations

Correspondence analysis is used to find possible associations between pattern identification proficiency and the query response spectrum: the associations for Rationality and the associations for Intuition. We discovered that the signal (the fraction not due to noise) obtained by the SVD (singular value decomposition) of the contingency table is much weaker for Rationality than it is for Intuition. For Rationality, we find two associations (Fig. 5), and for Intuition we find three (Fig. 6).

8.3

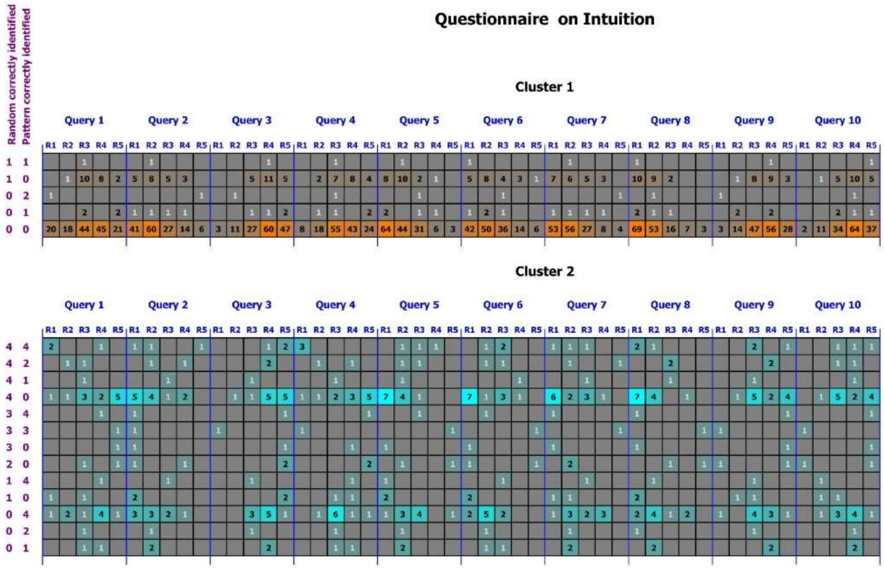


Fig. 5. The distribution of the nonzero entries of nonrandom and random pattern identification versus responses to the Intuition queries. Those rows that have no nonzero entries have been deleted. The brightness of the colors identifies the percentages of occurring entries for a response (this fraction is normalized according to the total number of respondents in a cluster). There are 3 respondents (1.7% of all 174) who correctly identified both the random patterns and the nonrandom patterns; they are distributed almost uniformly across all responses for all queries.

9 Discussion and Conclusion

The orientation in the world is based on heuristics that help the individuals navigate complex situations. Intuition and irrational beliefs can be products of perceptual and cognitive biases, as illusionary perception of patterns and becoming overly reliant on assigning meanings to these perceived patterns

There have been explanations proposed by cognitive scientists who infer the proliferation of supernatural beliefs to be the results of this and other cognitive biases (Johnson, 2009; Willard & Norenzayan, 2013). Furthermore, EMT also explains the existence of such biases as a compensation for potentially costly mistakes.

We decided to pursue the study of pattern perceptions by introducing two novel elements, namely the use of black and white patterns (rather than colored ones), and multiple (in our case: four) presentations by rotating and mirroring so as to assure that the chance of guessing the correct answer is minimized.

We emphasize that the study presented here is an example of a multidisciplinary approach. Here, psychological investigation techniques were combined with approaches based on novel technology and modern statistical techniques (including, but not limited to, dimension reduction and the use of clustering algorithms) in order to provide novel insights.

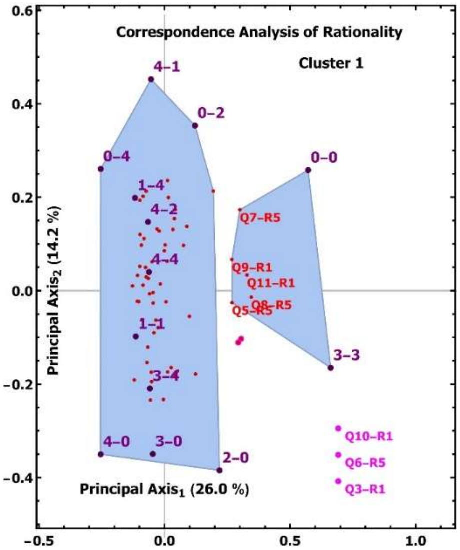


Fig. 6. The result of the correspondence analysis for pattern identification and Rationality: associations. The scree plot of squares of the singular values versus indices of the singular values identified only two singular values as non-noise. The reconstructed matrix has been graphed and the clusters identified via the clustering algorithm DBSCAN identified the categorical variables of nonzero query responses (red) that are associated with the pattern identification responses (purple). The fraction of the square of the Frobenius norm explained by the first two singular values is 40.2%. The magenta dots are query response choices that do not associate with any pattern identification pairs. The unlabeled query responses of one association have not been labelled, as the labels would have rendered that part of this graph unreadable. We point out that a strong association does not infer a high score on the response axes.

The dimensions for Rationality (60) and Intuition (50), combined with the those for pattern identification outcomes (26), make the multivariate analysis not only nontrivial, but require very large data sets for satisfactory noise minimization. Nonetheless, perhaps surprisingly, we found fascinating signals even in our relatively small data set. We are also aware of how difficult the test is for the participants because the presented patterns tested their skills quite severely. We were surprised that two clusters of the dimension-reduced feature vectors existed in both cases of thinking styles.
By using Artificial Neural Networks to cluster the participants based on their responses to the subscales of REIm (for Intuition and for Rationality, separately) and the results of the pattern identification task we identify two clusters for *each* subscale. In both cases we obtain one cluster in which the participants were more accurate in identifying the patters (or lack of) and another in which the participant exhibited a higher tendency for pareidolia and apoidolia. Each cluster showed associations with specific responses to specific queries, allowing us to estimate their similar psychological profile in the heuristic use in their daily life.

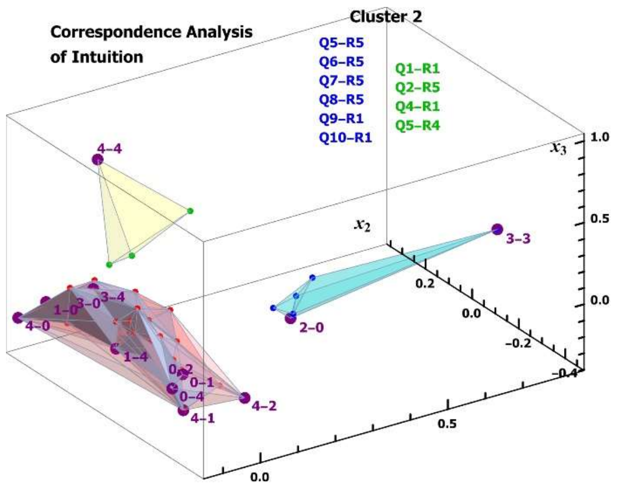


Fig. 7. The result of the correspondence analysis for pattern identification and Intuition: associations. The scree plot of squares of the singular values versus indices of the singular values identified three singular values as non-noise. The coordinates obtained by the reconstructed matrix have been graphed and the clusters were identified via the clustering algorithm DBSCAN. The identified categorical variables of nonzero query responses (green/blue/red) that are associated with the pattern identification responses (purple) show the associations: there are three associations. The fraction of the square of the Frobenius norm explained by the first two singular values is 58.1%. Most remarkable is the observation that the 'perfect' pattern identifiers (participants) associate with the query responses R1 for query Q1, R3 for query Q2, R1 for query Q4, and R4 for query Q5.

We also succeeded in demonstrating that these associations are the statistically justified descriptors (rather than correlations would be) and we found that there were more than one association (two for Rationality and three for Intuition) for those clusters of participants who were more successful in identifying patterns and distinguishing randomness and non-randomness (and therefore not experiencing pareidolia or apoidolia).
We note that the number of associations depended on the thinking styles and their statistical signal strength varied with thinking style. The results are promising; they mark an introduction of a novel approach to research of the phenomenon of apoidolia and extension of the existing research on pareidolia in a healthy population. The knowledge of how individuals orient themselves in the world based on their *believed* pattern presence can improve the design of future real world and virtual world (user) experiences. In the second case, this should be of utmost importance since the creator has almost absolute control over the environment presented to the user

Ethics Statement

The project was evaluated and approved by the Ethical Committee of the Faculty of Science, Charles University, as part of broader project (7/2018). GDPR regulations were followed at all times.

Funding and Acknowledgement

The research and data collection was supported by Charles University Grant Agency (GAUK "Cognitive Bias, Agency Detection and Pattern Recognition: An Intercultural Quantitative and Experimental Study of Supernatural Belief", project No. 1404120).
We thank to Ladislav Cupa for help with sample collection

Conflicts of Interest

The authors declare no conflict of interest.

References

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences* 4(1), 29–34.

Beh, E. & Lombardo, R. (2014). *Correspondence Analysis. Theory, Practice and New Strategies*. Chichester: Wiley. ISBN 978-1-119-95324-1.

Bellman, R.E. (1961). *Adaptive control processes: a guided tour*. Princeton University Press. ISBN 9780691079011.

Benzécri, J.-P. (1973). L'Analyse des Données. Volume II. L'Analyse des Correspondances. Paris, France: Dunod.

Blalock Jr, H.M. (1960). Social statistics. McGraw Hill, USA.

Boschetti, S., Binter, J. & Prossinger, H. (*in press*) Apoidolia: A New Psychological Phenomenon Detected by Pattern Creation with Image Processing Software together with Dirichlet Distributions and Confusion Matrices. *Proceedings in Adaptation, Learning and Optimization.*

Bressan, P. (2002). The connection between random sequences, everyday coincidences, and belief in the paranormal. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 16(1), 17–34.

Ester, M., Kriegel, H-P., Sander, J, Xu, X. (1996). In: Simoudis, E., Han, J., Fayyad, U. M. (eds.). *A density-based algorithm for discovering clusters in large spatial databases with noise*. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231. CiteSeerX 10.1.1.121.9220. ISBN 1-57735-004-9.

Fisher, R.A. (1956). Mathematics of a lady tasting tea. *The world of mathematics* III (part 8), pp. 1514–1521. Simon & Schuster, USA.

Greenacre, Michael (2007). *Correspondence Analysis in Practice*. Boca Raton: CRC Press. ISBN 9781584886167.

Johnson, D.D. (2009). The error of God: Error management theory, religion, and the evolution of cooperation. In: *Games,*

Groups, and the Global Good (pp. 169–180). Springer, Berlin, Heidelberg.
Murphy, K.P. (2012) Machine Learning. A Probabilistic Perspective. Cambridge, MA: The MIT Press.
Norris, P., & Epstein, S. (2011). An experiential thinking style: Its facets and relations with objective and subjective criterion measures. *Journal of Personality* 79(5), 1043–1080.
Tsao, D.Y. & Livingstone, M.S. (2008). Mechanisms of face perception. *Annual Rev Neurosci* 31, 411–437.
Wada Y., Kitagawa N., Noguchi K. (2003) Audio-visual integration in temporal perception. *Int J Psychophysiol* 50(1–2), 117–124. Doi: 10.1016/S0167-8760(03)00128-4.
Willard, A.K. & Norenzayan, A. (2013). Cognitive biases explain religious belief, paranormal belief, and belief in life's purpose. *Cognition* 129(2), 379–391.
Zhou, L.F., & Meng, M. (2020). Do you see the "face"? Individual differences in face pareidolia. *Journal of Pacific Rim Psychology* 14, e2.

Appendix

Table A-2. The queries for assessment of the two thinking styles. The responses are choices on a five-point scale from "Completely false for me" (R1) to "Completely true for me" (R5).

Query	Rationality	Intuition
Q1	I enjoy problems that require hard thinking.	I like to rely on my intuitive impressions.
Q2	I am not very good in solving problems that require careful logical analysis.	I often go by my instincts when deciding on a course of action.
Q3	I enjoy intellectual challenges.	I don't think it is a good idea to rely on one's intuition for important decisions.
Q4	I prefer complex to simple problems.	I trust my initial feelings about people.
Q5	I don't like to have to do a lot of thinking.	I tend to use my heart as a guide for my actions.
Q6	Reasoning things out carefully is not one of my strong points.	I enjoy learning by doing something, instead of figuring it out first.
Q7	I am not a very analytical thinker.	I can often tell how people feel without them having to say anything.
Q8	I try to avoid situations that require thinking in depth about something.	I generally don't depend on my feelings to help me make decisions.
Q9	I am much better at figuring things out logically than most people.	For me, descriptions of actual people's experiences are more convincing than discussions about "facts".
Q10	I have a logical mind.	I'm not a very spontaneous person.
Q11	Using logic usually works well for me in figuring out problems in my life.	
Q12	Knowing the answer without understanding the reasoning behind it is good enough for me.	

HCII RAINBOW

From Stone Age to New Age Statistics: How Neural Networks Overcome the Irreproducibility Problems in Choice Based Profile Creation

Hermann Prossinger[1,2][0000-0002-1114-7687], Silvia Boschetti[3][0000-0002-8048-4062], Daniel Říha[2][0000-0001-5142-4485], Libor Pacovský[2][0000-0000-0000-0000], and Jakub Binter[2][0000-0001-5304-2130]

[1] Department of Evolutionary Anthropology, Faculty of Life Sciences, University of Vienna, Vienna, Austria
[2] Faculty of Social and Economic Studies, University of Jan Evangelista Purkyně, Ústí nad Labem, Czech Republic
[3] Faculty of Science, Charles University, Prague, Czech Republic
hermann.prossinger@univie.a.at

Abstract. In the gaming world, as well as in assistive technologies, profile creation and further characterizations of humans are basic. Oftentimes, Likert-type scales are needed to collect responses in questionnaires. A subsequent mistake by the analyst occurs when he/she computes the composition score derived from these response categories. Because: such a score can only be computed if these categorical responses are converted into computable numbers. An AI method, namely the application of artificial neural networks, can extract information and overcome various erroneous statistical methods (such as: linearizing responses of Likert scales, computing scores and correlations, and disregarding the 'curse of dimensionality'). We collected data from 480 respondents who were asked to specify boundaries between the colors of the blackbody spectrum ('rainbow'). We first used an auto-encoder for dimension reduction, then searched for categories by implementing clustering algorithms, computed likelihood plots, and calculated confusion matrices based on Dirichlet distributions.

We found that every respondent was a member of one of only three clusters. Each cluster is characterized by a different distribution of the color boundaries between purple-blue, blue-green, and so on. Surprisingly, the boundaries between some colors in one cluster where within the color interval for members of another cluster. Conclusion: where people see a color boundary is far from obvious even within a shared range.

There are implications for several fields that have been using 'Stone Age statistics' involving scores, etc. as listed above, among them psychology, sociology, behavioral economy, and human-computer interaction. The modern approach, based on AI, should be adopted by researchers in these fields to ensure reproducibility and provide insight into participant/user/citizen profiles.

Keywords: One-hot Encoding, Artificial Neural Network, Confusion Matrix, Visible Spectrum, Color Boundaries, Kernel Density Estimation, Profile Creation, Cold Start Problem.

10 Introduction

Currently, profile creation is a big topic in many fields involving technology. In an ideal case, there exists both a large data set and prolonged development of the field that has already been supplied with 'natural' layperson terminology for distinguishing features defined by the layperson's limited choices. Thus, while this may not be the case in many fields such as technology users and online gamers, for citizens responding to nudges, etc. that are uninitiated, such approaches cannot be applied from scratch. This brings the *cold start problem* into play (Eke et al., 2019). The system that is used for recommendation, profiling and divergence is unable to provide any meaningful inferences since too much information is lacking (Eke et al., 2019).

There are two main approaches to deal with this described problem. One involves extended, long-term data collection, including updates and adjustments, and a bean-counting summary description; the second one bases the distinctions on the data gathered from the prospective users via a questionnaire-type set of queries (Eke et al., 2019).

Questionnaires, the major tool in political polls, communication research, customer satisfaction research, and psychometrics, consist of a list of queries. Participants are requested to respond by making a choice for each of the categorical variables offered. For some queries, the options can be ranked; then the response option list is called a Likert scale (for review see Jebb et al., 2021). The length of the list of response options need not be the same for every query, nor need all queries be Likert scales. Indeed, in the study we present here, we specifically cannot ensure the same number of response options for every query.

Table 3. A selection of possible conversions (maps) of categorical variables that are responses to queries.

Label	Map	A	B	C	D	E	F
Conv	Conventional	1	2	3	4	5	6
Form	Formula I	8	10	12	15	18	25
MGP	MotoGP	10	11	13	16	20	25
Alp	FIS Alpine	47	51	55	60	80	100
Pri	Odd prime numbers	3	5	7	11	13	17
Lo1	Lottery 1	6	10	12	13	15	41
Lo2	Lottery 2	2	9	11	26	39	42
RConv	Reciprocal Conv	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1
RPri	Reciprocal Odd Primes	$\frac{1}{17}$	$\frac{1}{13}$	$\frac{1}{11}$	$\frac{1}{7}$	$\frac{1}{5}$	$\frac{1}{3}$

The conventional approach to analyzing a questionnaire with these response distributions is to map the categorical responses into cardinal numbers. Doing so in the conventional manner leads to individuals who answered differently (for example, 1, 5, 3 by one individual and 3, 4, 2 by another to 3 questionnaire queries with 5 response options per query) to be assigned the same profile (9 and 9) — a statistical bias that will be included in — and render meaningless — the subsequent analyses. Even if the response options are offered to the participant as numbers, these numbers are by no means cardinal numbers (they are ordinal numbers; Blalock, 1960), so it is impossible to calculate variances, averages or other statistical point estimators. The response options (categorical variable registrations) must be converted into cardinal numbers. This mapping of ordinal numbers into cardinal numbers is by no means unique. Table 1 shows a selection of the infinitely many maps that are possible.

The conventional mapping (Conv in Table 1) is the most ubiquitous. This choice of mapping has been criticized (Prossinger et al., 2022). Nonetheless, researchers in the aforementioned fields still use it almost exclusively. To summarize: the variances and means of the responses in a questionnaire will vary, depending on the mapping chosen.

To avoid specific, case-study implications and maintain the linear nature of the presented phenomenon, we decided to present our participants with the blackbody spectrum ('rainbow') of an object at 6000 K surface temperature and asked them to choose the boundaries separating two neighboring colors. The human eye is very perceptive to these wavelengths (as these are the ones present in large amplitudes in sunlight). These colors have been previously investigated in the classical work about human visual perception *Wavelength discrimination at detection threshold* by Mullen and Kulikowski (1990). More recently, Vlad et al. (2021) adopted a similar approach but used digitized color production to ensure the mechanism is applicable to digital media (the ubiquitous current devices). Our aim was to map how many clusters (profiles that share commonalities) will form in such a straightforward task as color boundary perception. Another one was to identify the existence of overlaps between the clusters based on participant's visual perceptions. This paper should be mainly understood as a methodological study of *hot start profiling*.

11 Materials

11.1 Questionnaire

A spectrum of visible light (the blackbody spectrum of light emitted by a black body at 6000 K surface temperature) was presented. At certain positions along the spectrum, letters of the Latin alphabet (hence ordered categorical variables) were supplied (Fig. 1). Each participant was asked to choose the letter he/she considered closest to the boundary queried (such as blue-green). In total, a total of five boundaries were queried: purple-blue, blue-green, green-yellow, yellow-orange, and orange-red. Each participant therefore entered a total of five responses (as letters) into the data set. Query responses were collected with Qualtrix®.

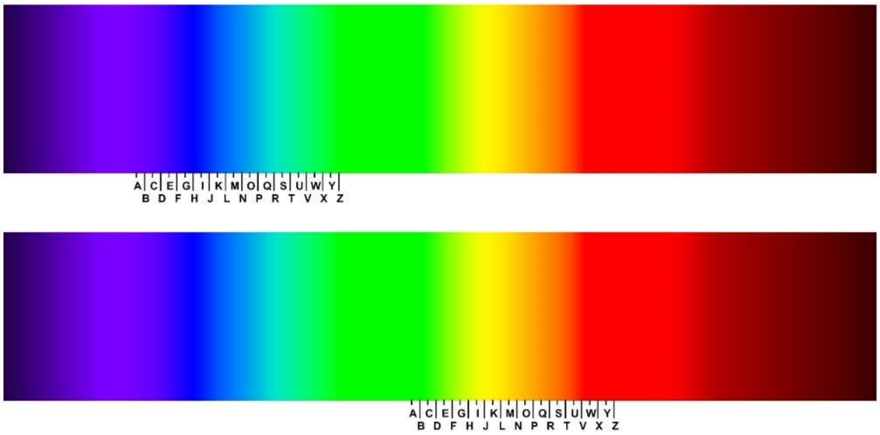


Fig. 8. The visible spectrum together with the two category markings presented to the participants. The participants were to choose the letter (category) that they considered closest to their perception of the boundary between the colors queried in the questionnaire. Although the marked rulings appear with constant intervals, it is not clear (to the participant) whether the wavelengths corresponding to these markings are also equidistant. Above: The spectrum presented for the queried boundaries purple-blue and blue-green. Below: The spectrum presented for the queried boundaries green-yellow, yellow-orange and orange-red.

11.2 Participants

A total of 480 participants were involved in this study; each participant supplied age and biological sex. Their ages ranged from 18–50 years for both the 170 males and 310 females. We estimated the age distributions and their descriptive parameters using a KDE (Kernel Density Estimation) with a Gaussian kernel

12 Methods

12.1 Age Distributions, male versus female

We used KDE to find the distribution of ages of the female and the male raters, separately. We then determined the estimators, in particular the $HDI_{95\%}$ (the Highest Density Interval at 95% significance; Kruschke, 2015). A large overlap of this interval documents no significant difference in the age distributions.

12.2 Feature vector construction: One-hot encoding

Each score by each participant for each queried color boundary is an ordinal variable. Contrary to the fallacious method of mapping these scores into cardinal integers, we convert the sequence of scores by a participant into a (concatenated) feature vector. For example, assume the score for a participant (in this example, we use the choices from participant #13) for the purple-blue boundary is E (from the range of all registered scores A–P (Fig. 1) ⟶ 16 options), the score for the blue-green boundary is S (from the range of all registered scores K–W (Fig. 1) ⟶ 13 options), the score for the green-yellow boundary is H (from the range of all registered scores C–K (Fig. 1) ⟶ 9 options), the score for the yellow-orange boundary is L (from the range of all registered scores H–T (Fig. 1) ⟶ 13 options), and, the score for the orange-red boundary is S (from the range of all registered scores M–Z (Fig. 1) ⟶ 14 options). Then the one-hot encoded feature vectors are

for purple-blue: $(0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)^T$

for blue-green: $(0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0)^T$

for green-yellow: $(1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)^T$

for yellow-orange: $(0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)^T$

for orange-red: $(0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0)^T$.

The one-hot encoded feature vector for this participant is then

$$(0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \cdots 0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0)^T.$$

The feature vector for every participant's five scores has $16 + 13 + 9 + 13 + 14 = 65$ components, and the norm of this feature vector is $\sqrt{5}$. In the case of participant #13, the nonzero entries are the components 5, $16 + 9 = 25$, $16 + 13 + 1 = 30$, $16 + 13 + 9 + 5 = 43$, and $16 + 13 + 9 + 13 + 7 = 58$.

12.3 Dimension Reduction and Clustering

Each participant's feature vector is therefore a point in 65-dimensional space. There are three reasons why a dimension reduction algorithm is needed: (a) not all possible feature vectors occur, (b) due to the 'curse of dimensionality' (Bellmann, 1961), the variance grows (it is additive and always positive) with the squares of the realizations of the random variable(s) — therefore much faster than the signal, and (c) the feature vectors of the participants have interdependencies (which we are looking for).
There are two modern methods that can be used for dimension reduction: SVD (singular value decomposition) and a ANN (artificial neural network). In SVD, we look for a linear interdependence of the feature vectors, while for ANN, nonlinear interdependencies can also be included. We choose a special ANN, namely an autoencoder, in order to reduce each participant's feature vector to a 2D one (Fig. 4a).
The dimension-reduced feature vectors are not uniformly distributed in the plane. We use the DBSCAN clustering algorithm (Ester et al., 1996) to detect clusters. In order to determine how well the clusters are separated, we use a KDE distribution with a triweight kernel.
We construct the confusion matrix to estimate the significance of the overlap. For each pair of clusters (cluster_A and cluster_B, say), we compute the confusion matrix

$$\begin{pmatrix} \dfrac{pdf(KDE_{\text{cluster}_A}) > pdf(KDE_{\text{cluster}_B})}{n_{\text{cluster}_A}} & \dfrac{pdf(KDE_{\text{cluster}_A}) \le pdf(KDE_{\text{cluster}_B})}{n_{\text{cluster}_A}} \\ \dfrac{pdf(KDE_{\text{cluster}_B}) \le pdf(KDE_{\text{cluster}_A})}{n_{\text{cluster}_B}} & \dfrac{pdf(KDE_{\text{cluster}_B}) > pdf(KDE_{\text{cluster}_A})}{n_{\text{cluster}_B}} \end{pmatrix}$$

We 'backtrack' from the dimension-reduced feature vectors to the scores for all participants in a cluster.
In order to determine the significance of the differences in the five boundaries for the different clusters, we again use the machinery of confusion matrices. We determine the union of entries, for a given boundary and a given pair of clusters, and tally the frequencies. The *pdf*s of these frequencies are the concentration parameters of two Dirichlet distributions, one for each cluster of the clusters being compared.

13 Results

Fig. 2 and Table 2 show that the distributions of the male and female raters have different modes. We use KDE (kernel density estimation) with a Gaussian kernel because, as Fig. 1 shows, it is not to be expected that raters of either sex have a parametric distribution or even a superposition of one or two such parametric distributions. We also note that modes and expectation values differ between the sexes and also for the same sex. The $\text{HDI}_{95\%}$ uncertainty interval is very broad, so we can consider the distributions for both sexes to be comparable to a uniform distribution of respective ages. The confusion matrix shows that the two distributions are not significantly different. There is, therefore, no age-effect for the boundary sets we find.

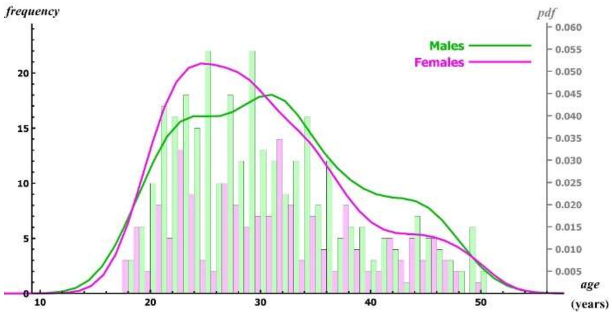


Fig. 2. The distribution of male and female ages of the participants. The histogram shows the ages entered by the participants. The curves are the *pdf*s (probability density functions) of the KDE (kernel density estimation) distributions of the ages, by biological sex, using a Gaussian kernel. The graphed *pdf* curves have been scaled (in this figure) so as to be comparable with the histogram rectangles. The numerical values of the *pdf*s are shown as a (gray) scale on the right. The age distributions of the males and females are not significantly different ($significance = 55\%$).

Table 4. The descriptors of the ages of the participants, separated by (biological) sex. Only the mean is a point estimator. All other descriptors are derived from the KDE (using a Gaussian kernel). N is the sample size and $\mathbb{E}$ is the expectation value. Because the KDE is neither a symmetric nor a parametric distribution, calculating a standard deviation is not meaningful. We note that the uncertainty $HDI_{95\%}$ is neither symmetric about the mode, nor about the expectation value. We also note that the modes cannot be calculated from the raw data, but must be estimated from the KDEs. The distributions (KDEs), which have been estimated from the data, are not significantly different (see text). Consequently, none of the descriptors (except for the sample sizes) are significantly different.

Age Descriptor	Male	Female
N	170	310
Mean (years)	31.4	30.2
$\mathbb{E}$ (years)	30.2	29.1
$HDI_{95\%}$ (years)	16.2–49.8	16.8–48.8
Mode (years)	31.0	24.6

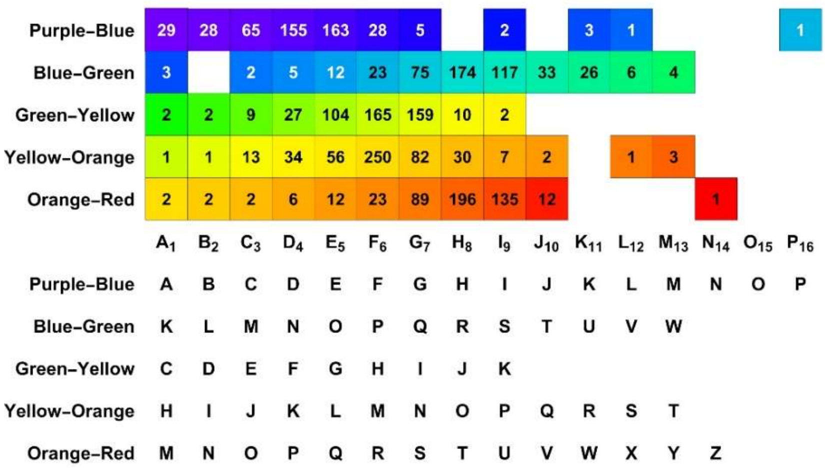


Fig. 3. The grid of tallies of boundary choices made by the 480 participants. The numbers within the squares are the number of participants that chose the (displayed) color as the boundary. Each horizontal row of tallies sums to 480. The colors of the squares are the colors of the boundary chosen by the participants. The letters chosen by the participants differed from boundary query to boundary query. These letters have been replaced by indexed symbols in the first row. Below this row of indexed symbols are the actual labels (Fig. 1) of the participant's scores. There is no encoding in the choice of black or white numbers of the tallies displayed; rather, the choice of black or white was made to enhance the contrast between the displayed numbers and the background color.

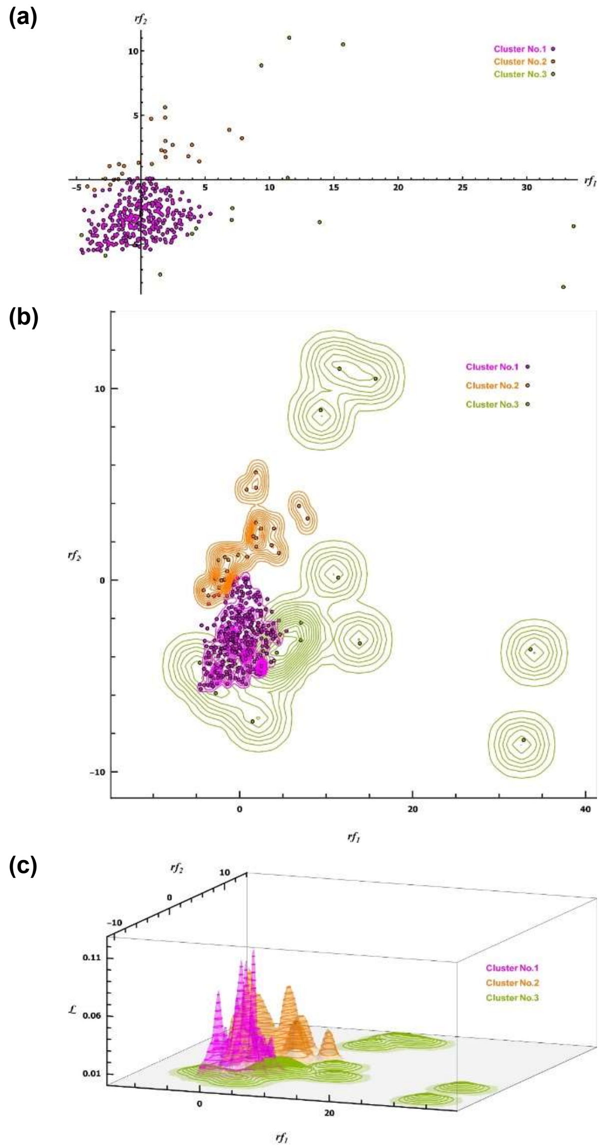


Fig. 4. The distributions of clusters of the dimension-reduced feature vectors of the color boundaries chosen by the participants. **(a)** The distributions of each cluster in the dimension-reduced feature vector space (here: a plane). The separation of clusters seems not to be very large in some regions of the 2-D feature space. **(b)** The projection of the contour plots of the likelihood surfaces obtained by the KDEs (each of the three with a triweight kernel of the functional form $\frac{35}{32}(1-u^2)^3$). Contours for each likelihood surface are in steps of $\frac{1}{15}\mathcal{L}_{\max}$ of the cluster ($\mathcal{L}$ is the likelihood). Some contours overlap. **(c)** A 3D graph of the likelihood surfaces obtained by the KDEs. The likelihood surfaces are very broad for Cluster No.3 and very peaked for Cluster No.1. As a consequence, the separation between clusters is highly significant (Table 3).

Fig. 3 shows the heat map of the boundary choices. We observe that the distribution of boundary choices varies considerably for some color boundaries. We note that: (a) Only one boundary query has no gaps, while the others have one, three, and five. (b) The 'length' (number of scores/responses) varies from query to query. It is therefore impossible to use Cronbach's Alpha or some other coefficient of reliability. We also note that the largest tally numbers for each boundary do not match a central tendency for the boundary query.

Table 5. The confusion matrices of significant overlap between the likelihood functions estimated via KDE using a triweight kernel (Fig 4b–4c). All the overlaps are significantly different. The top row in each confusion matrix is for the cluster with the smaller ordinal number (index).

Cluster No.1 ⟷ Cluster No. 2	$\begin{pmatrix} 98.64 & 1.36 \\ 0 & 100 \end{pmatrix}$ %

Cluster No.1 ⟷ Cluster No. 3	$\begin{pmatrix} 98.87 & 1.13 \\ 0 & 100 \end{pmatrix}$ %
Cluster No.2 ⟷ Cluster No. 3	$\begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}$ %

We obtain three clusters of boundary sets (Fig. 4). The clusters were found using the DBSCAN algorithm (Ester et al., 1996). One cluster is by far the largest, with 441 participants (91.9%), the second largest cluster consists of 25 participants (5.2%), and the smallest cluster consists of 14 participants (2.9%).
The pairwise overlap between the likelihood functions is so small that the confusion matrices are either diagonal or close to diagonal (Table 3). We therefore conclude there is no significant overlap and the clusters are significantly different.

Table 6. The confusion matrices of significant differences between the boundaries in the different clusters. The entries in these confusion matrices are in %. These confusion matrices have been calculated by generating 15000 random numbers from each of the Dirichlet distributions, boundary by boundary. The top row in each confusion matrix is for the cluster with the smaller ordinal number (index). We observe that only for Cluster 2 versus Cluster 3 are the off-diagonal elements of the confusion matrices nonzero. However, even for this cluster comparison, the difference is significant.

Boundary	Clusters 1⟷2	Clusters 1⟷3	Clusters 2⟷3
purple⟷blue	$\begin{pmatrix} 100.0 & 0.0 \\ 0.0 & 100.0 \end{pmatrix}$	$\begin{pmatrix} 100.0 & 0.0 \\ 0.0 & 100.0 \end{pmatrix}$	$\begin{pmatrix} 99.90 & 0.10 \\ 0.10 & 99.90 \end{pmatrix}$
blue⟷green	$\begin{pmatrix} 100.0 & 0.0 \\ 0.0 & 100.0 \end{pmatrix}$	$\begin{pmatrix} 100.0 & 0.0 \\ 0.0 & 100.0 \end{pmatrix}$	$\begin{pmatrix} 99.59 & 0.41 \\ 0.35 & 99.65 \end{pmatrix}$
green⟷yellow	$\begin{pmatrix} 100.0 & 0.0 \\ 0.0 & 100.0 \end{pmatrix}$	$\begin{pmatrix} 100.0 & 0.0 \\ 0.0 & 100.0 \end{pmatrix}$	$\begin{pmatrix} 97.75 & 2.25 \\ 1.89 & 98.11 \end{pmatrix}$
yellow⟷orange	$\begin{pmatrix} 100.0 & 0.0 \\ 0.0 & 100.0 \end{pmatrix}$	$\begin{pmatrix} 100.0 & 0.0 \\ 0.0 & 100.0 \end{pmatrix}$	$\begin{pmatrix} 99.98 & 0.02 \\ 0.03 & 99.97 \end{pmatrix}$
orange⟷red	$\begin{pmatrix} 100.0 & 0.0 \\ 0.0 & 100.0 \end{pmatrix}$	$\begin{pmatrix} 100.0 & 0.0 \\ 0.0 & 100.0 \end{pmatrix}$	$\begin{pmatrix} 99.53 & 0.47 \\ 0.41 & 99.59 \end{pmatrix}$

Visual inspection of the boundary distributions for the different clusters reveals no evident differences (Fig. 5). This is due to the fact that the human visual system is not good at detecting non-linear interdependencies. The ANN detects interdependencies of the chosen boundaries that we are surprised at observing when confronted with the graph.

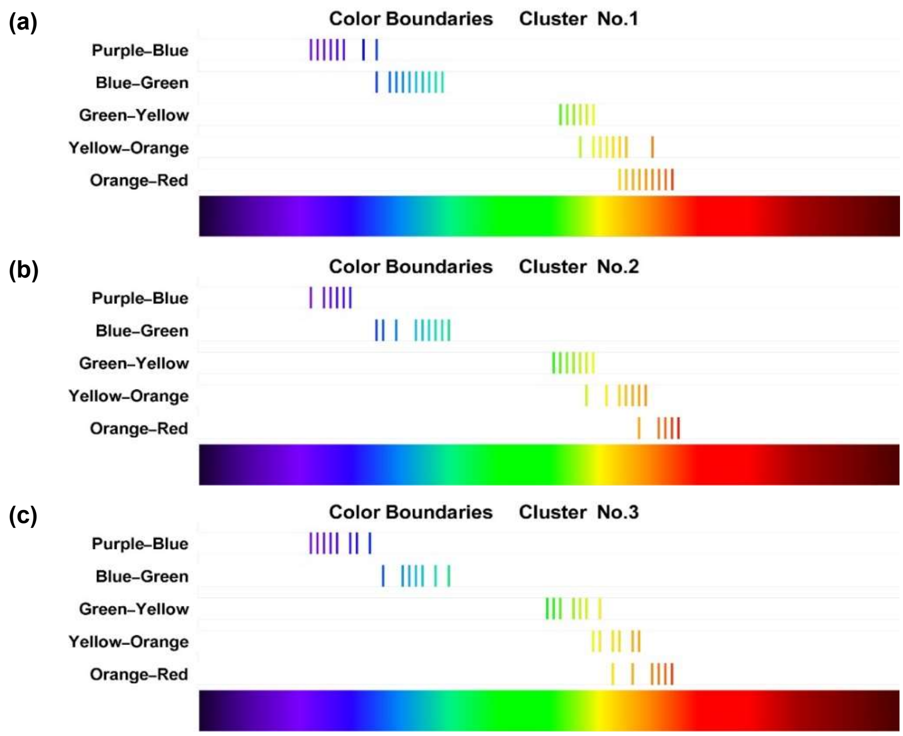


Fig. 5. The distributions of boundaries chosen by the participants in the three clusters. The multiplicity of the chosen boundaries is not shown. **(a)** Cluster No. 1 (441 participants; 92%), **(b)** Cluster No. 2 (25 participants; 5,2%), and **(c)** Cluster No. 3 (14 participants; 2.9%).

14 Discussion and Conclusions

Profile creation is becoming more and more important as the method availability is quasi-exponentially growing because of technological advancements. The *cold start problem* can be resolved by obtaining the data from the prospective users beforehand and creating groups (clusters) based on their responses using questionnaire queries. Many 'antiquated' techniques have been implemented to deal with this problem: from Cronbach´s Alpha and MacDonald's Omega to various versions of FA (factor analysis), and to SEM (structural equation modeling). Because both FA and SEM lack a rigorous mathematical foundation (when using ordinal responses), rules of thumb are employed to achieve a 'presentable' result and satisfy peer reviewers. Unfortunately, FA and SEM of ordinal responses are still in widespread use in the psychological and behavioral sciences (thereby producing non-repeatable and biased outcomes).

AI (artificial intelligence) has become a buzzword. However, in an actual study such as this one, which focuses on the (seemingly simple) task of finding color boundaries by applying a novel methodology on previously tested outcomes, we were able to identify three independent clusters — an unanticipated new finding. To our knowledge, no previously tested outcomes have found any clustering. One cluster constitutes over 90% of the participants and other methods would have failed to identify the remaining two clusters that are decisively different (and small). Since optical perception mechanisms, like color perception, are thought to be non-uniformly distributed in the human population, this finding of the existence of three clusters is an important insight because, among other implications, it signals a warning to avoid FA and SEM.

We show how the tools of AI can be implemented, how they can be applied to categorical variables (here: ordinal ones), and how dimension reduction methods can overcome the 'curse of dimensionality' (Bellman; 1961). We show how AI algorithm outputs can then be analyzed by using clustering algorithms, then computing likelihood plots, and finally calculating confusion matrices; all these allow for finding and defining categories of the participants as profiles. Many of these methods are implementable in both supervised AI and unsupervised AI.

It should be pointed out that the task was intentionally chosen to be straightforward and would avoid distractors such as emotional involvement, political preferences or socio-historical components. We were still able to identify decisively different groups (clusters, in statistical parlance) in a population. If we take into account the above mentioned possibilities, the proportions in the clusters can be expected to exhibit the different clusters in a population. Bearing in mind that governments (nudging) or corporate companies (human-computer interaction) may involve millions of individuals being affected by their decisions and their profiling, these new methods should be adopted as soon as possible to properly address the complexity within the population and provide meaningful solutions. These approaches should by all means be combined with the prospective adaptations based on user/citizen profiles to achieve best results (Eke et al., 2019; Farnandi et al., 2018).

Ethics Statement

The project was evaluated and approved by the Ethical Committee of the Faculty of Science, Charles University, as part of broader project (7/2018). GDPR regulations were followed at all times.

Funding and Acknowledgement

This article was authored with the support of the Doctoral School: Applied and Behavioral Studies project (Grant Number: CZ.02.2.69/0.0/0.0/16_018/000272), University of Jan Evangelista Purkyně, Ústí nad Labem, Czech Republic. D.Ř. is further funded by the Ministry of Education, Youth and Sports, Czech Republic and the Institutional Support for Long-term Development of Research Organizations, Faculty of Humanities, Charles University, Czech Republic (Grant COOPERATIO "Arts and Culture").

Conflicts of Interest

The authors declare no conflict of interest.

References

Bellman, R.E. (1961). *Adaptive control processes: a guided tour*. Princeton University Press. ISBN 9780691079011.

Blalock Jr, H.M. (1960). *Social statistics*. McGraw Hill, USA.

Eke, C.I., Norman, A.A., Shuib, L., & Nweke, H.F. (2019). A survey of user profiling: State-of-the-art, challenges, and solutions. IEEE Access, 7, 144907–144924.

Ester, M., Kriegel, H-P., Sander, J, Xu, X. (1996). In: Simoudis, E., Han, J., Fayyad, U.M. (eds.). *A density-based algorithm for discovering clusters in large spatial databases with noise*. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231. CiteSeerX 10.1.1.121.9220. ISBN 1-57735-004-9.

Farnadi, G., Tang, J., De Cock, M., Moens, M. F. (2018). User profiling through deep multimodal fusion. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (pp. 171–179).

Jebb, A. T., Ng, V., Tay, L. (2021). A review of key Likert scale development advances: 1995–2019. *Frontiers in Psychology* 12, 637547.

Kruschke, J.K. (2015) *Doing Bayesian Data Analysis*. *A Tutorial with* R, JAGS, *and* STAN. Elsevier, Waltham, MA, US A

Mullen, K.T., & Kulikowski, J.J. (1990). Wavelength discrimination at detection threshold. JOSA A 7(4), 733–742.

Prossinger, H., Binter, J., Machova, K., Riha, D., Boschetti, S. (2022). Machine Learning Detects Pairwise Associations between SOI and BIS/BAS Subscales, making Correlation Analyses Obsolete. In: *Human Interaction & Emerging Technologies* (IHIET-AI 2022). *Artificial Intelligence & Future Applications*. Vol 23. AHFE International, USA. doi: 10.54941/ahfe100903.

Vlad, V., Toti, M., Dumitru, S., Simota, C., Dumitru, M. (2021). Developing reliably distinguishable color schemes for legends of natural resource taxonomy-based maps. Cartography and Geographic Information Science, 48(5), 393-416.

technologies

MDPI

Article

Determination of "Neutral"–"Pain", "Neutral"–"Pleasure", and "Pleasure"–"Pain" Affective State Distances by Using AI Image Analysis of Facial Expressions

Hermann Prossinger [1,*], Tomáš Hladký [2], Silvia Boschetti [2,3], Daniel Říha [2,3] and Jakub Binter [2,3]

1 Department of Evolutionary Biology, Faculty of Life Sciences, University of Vienna, A-1030 Vienna, Austria
2 Faculty of Humanities, Charles University, 182 00 Prague, Czech Republic; tomas.hladky@fhs.cuni.cz (T.H.); silvia.boschetti@natur.cuni.cz (S.B.); daniel.riha@fhs.cuni.cz (D.Ř.); jakub.binter@fhs.cuni.cz (J.B.)
3 Faculty of Science, Charles University, 128 00 Prague, Czech Republic
* Correspondence: hermann.prossinger@univie.ac.at

Abstract: (1) Background: In addition to verbalizations, facial expressions advertise one's affective state. There is an ongoing debate concerning the communicative value of the facial expressions of pain and of pleasure, and to what extent humans can distinguish between these. We introduce a novel method of analysis by replacing human ratings with outputs from image analysis software. (2) Methods: We use image analysis software to extract feature vectors of the facial expressions neutral, pain, and pleasure displayed by 20 actresses. We dimension-reduced these feature vectors, used singular value decomposition to eliminate noise, and then used hierarchical agglomerative clustering to detect patterns. (3) Results: The vector norms for pain–pleasure were rarely less than the distances pain–neutral and pleasure–neutral. The pain–pleasure distances were Weibull-distributed and noise contributed 10% to the signal. The noise-free distances clustered in four clusters and two isolates. (4) Conclusions: AI methods of image recognition are superior to human abilities in distinguishing between facial expressions of pain and pleasure. Statistical methods and hierarchical clustering offer possible explanations as to why humans fail. The reliability of commercial software, which attempts to identify facial expressions of affective states, can be improved by using the results of our analyses.

Keywords: image processing; artificial intelligence; facial expressions; affective state expression; facial pain expression; facial pleasure expression; BDSM videos; hierarchical agglomerative clustering; autoencoder neural network

Citation: Prossinger, H.; Hladký, T.; Boschetti, S.; Říha, D.; Binter, J. Determination of "Neutral"–"Pain", "Neutral"–"Pleasure", and "Pleasure"–"Pain" Affective State Distances by Using AI Image Analysis of Facial Expressions. *Technologies* **2022**, *10*, 75. https://doi.org/10.3390/technologies10040075

Academic Editors: Vijayakumar Varadarajan and Mario Munoz-Organero

Received: 30 March 2022
Accepted: 18 June 2022
Published: 22 June 2022

1. Introduction

This manuscript reports results that are extensions of the pilot study published in 2021 [1]. There, we investigated, using AI, whether it is possible to distinguish the facial expression of pain from that of pleasure in women.

Consider the following scenario: two persons are having a face-to-face conversation (rather than telephoning). In addition to the exchange of words, the conversants unavoidably register each other's facial expressions. Whether these facial expressions convey and, if yes, how much information is the starting point of the research presented in this paper.

The information transmitted is the superposition of three contributions: acoustic, facial, and semantic. Anyone who attempts to decompose this superposition is faced with extraordinary challenges. A promising approach would be to devise a laboratory or field work setting wherein actors convey their affective states of pain and of pleasure (hopefully convincingly) to an audience. By turning off the microphone or erasing the soundtrack, we are in the position to analyze what is being solely transmitted via facial expressions.

Aim: we wish to show how the use of AI technology that analyzes images is superior to humans' abilities to solve these same facial expression decoding challenges.

1.1. Overall Benefits of the Insights We Present in This Manuscript

Due to human evolution, the ability to produce facial expressions does not infer the ability to perceive them. In fact, the shortcoming in this lack of inference is the perception part. We can gain insights by overcoming this shortcoming by relying on technology as an alternative. The result(s) obtained by image analysis software does not, as we will show, reproduce the human result. It seems that computer vision is not compromised by a human brain's visual perception limits. Therefore, we argue that the insights gained are beneficial to our understanding and appreciation of visual perception mechanisms.

1.2. Using AI as a Novel Approach to Analyzing Facial Expressions of Pain and Pleasure

One way to perform this analysis would be to query the members of an audience as to their evaluation of presented facial expressions. Another approach is to analyze the expressions using image analysis software. We present research results of this latter approach in this paper.

There exists a considerable amount of literature [2–6] dealing with the recognition and identification of affective states putatively communicated by facial expressions. Novel in the approach we present here are the following: (1) to use one AI tool to extract faces from video frames; (2) to use another AI tool (an autoencoder) to determine their dimension-reduced feature vectors; (3) to quantify their differences (if any); and (4) to implement the machinery of multivariate statistical analysis, combined with clustering algorithms, to detect noise contributions and identify patterns.

For technical and historical reasons, we limit ourselves, in this paper, to focusing on actresses while they express three affective states: pain, pleasure, and neutral.

1.3. Previous Reasearch into Facial Expression of (Intense) Affective States

While interpersonal interaction via verbal exchanges involves language, facial expressions are also, arguably, a further method to "advertise" one's affective state and even more so during many social interactions in which the acoustic channel is blocked (such as a pantomime or an interaction through a window or while observing from a distance).

For one person to estimate the affective state of some other person, his/her decoding ability is the foremost prerequisite. Further support for the hypothesis of universality comes from intercultural studies that have documented that expressions of emotion, among them joy, fear, sadness, anger, and surprise (but perhaps neither pain nor pleasure; both are not emotions), are to a large degree universally understood [7–9]; this is also supported by methodologies using AI [10,11].

Many expressions of affective states are closely linked to these in combination; indeed, there is a widespread consensus in psychology that there exist only a small number of emotions expressed and experienced solely. Some of these are also characterized by specific—and identifiable—facial muscle contractions [9].

This inferred existence of the in-between ones has led to a codification called FACS (Facial Action Coding System; [9]). It relies on so-called action units (AU); each unit is based on the contraction of a subset of facial muscles to create, in concerto, the intended facial expression. The derived methods are used both descriptively (in the analysis of the behavior of an individual) and also prescriptively as a template for the animation of facial expressions [12]. A study [13] of several animations that have been used in computer graphics focused on the distinction between the facial contribution and the perceived affective states associated with pain and with sexual climax (pleasure). That study succeeded in demonstrating a difference of mental representations within the onlooker of the two affective state expressions.

One can, of course, attempt to down-regulate the expressiveness and thereby control the amount of information shared (a poker face comes to mind) but there may be limits in the event of extreme experiences. In the context of sexuality, this issue of pain resembling grimace during sexual climax has already been described in the so called "Kinsey Report" (1953). Further examples involving other social situations have recently come to the fore [14].

Pain and pleasure, as signals of extremely important experiences in the lives of humans, should be sufficiently well-recognizable by social animals, which humans certainly are. There is an ongoing debate about the communicative value of these facial expressions [15]. One of the arguments is that they are essential because they provide the interacting partner with a signaling value that he/she can attribute correctly [16]. One of the counterarguments is that the groups of muscles contracted during the intense affective states of pain and pleasure are similar and therefore the only inferable information relates to intensity but does not provide any valence of the signal [13,17].

Thus, there is a possibility that, even if a difference is indeed present, human perception is not calibrated well enough to reliably and reproducibly distinguish such details. Perhaps human vision is unable to adequately cope with the noisiness of the signal and humans may perceive it as individual differences in expression. In essence, the research involving facial expressions perceives a limitation in how to overcome these difficulties in traditional field studies, arguably a gap in available methodology [6]. AI facial analysis should be able to identify such differences, thus bridging this gap and implying that it is superior to human vision in this regard. Already in our pilot study [1], this distinguishability was found in women but not in men. Based on these published outcomes, we increased the sample size but restricted the study to actresses only.

Facial expressions in some cases seem to be ambiguous to raters whenever they are not presented with further cues (such as vocalization or body posture); stimuli are oftentimes acted instead of being real. In one study [18], actors had been trained to act out fear and anger. The accuracy of their acting ability for the affective state of fear was tested by comparisons with real-life recordings; the acted expressions differed greatly, it was found, from the real-life recordings and were more difficult to recognize. In our attempt to further increase the demands of rigor regarding facial expressions of pain and of pleasure, we used real bondage, discipline, and sadomasochism (BDSM) acts performed by professional actors in commercially available videos.

The use of BDSM videos of private persons would infringe on their privacy rights and we could not reliably infer which affective state is being expressed (we would have to rely on their statements). In the case of BDSM videos with professional actresses, on the other hand, we can be certain that the displayed affective state is expressed (being monitored by the director during production) while the issue of privacy rights is moot. Even though the actresses are professionals and are probably anticipatory of what the consumer is seeking or expecting, the stimuli should be considered semi-naturalistic since we, the authors, intentionally chose companies and brands that have a reputation for their realism. There are indicators (such as bruises) that the scenario is not enacted symbolically but that the actresses' experiences are genuine.

1.4. Novelty of the Approach Presented in This Paper

While the analyses of facial expressions of pain and pleasure by BDSM actresses are, we claim, novel, per se, we go several steps further. We rated the facial expressions using image analysis software and other AI tools. We present outcomes that indicate to what extent image analysis software is more reliable than human ratings of the facial expressions of pain and pleasure.

We extracted feature vectors from the images and quantified differences between the facial expressions by calculating the Euclidean distance between the dimension-reduced feature vectors (thus defining the affective state distances). We used these distances to determine their maximum likelihood (ML) distributions, quantified the noise, and looked for patterns of these noise-free dimension-reduced feature vectors, which imply patterns in the actresses' facial expressions.

1.5. Fields of Study in Which the Results Are of Importance

Applications using image analysis software have become standard in the world of AI. However, because the facial expression of pleasure, for instance, is unreliable, there is a potential confusion with the facial expression of pain. The results we present here quantify the putative unreliability, which should contribute to the development of more refined algorithms in image analysis software attempting to distinguish these affective states.

2. Materials and Methods

2.1. Materials

We scanned 20 BDSM videos, each showing professional actresses in action. In each video chosen, using the development of the plot as a reference, five frames with almost format-filling faces—three frames displaying neutrality, one frame pain, and one frame pleasure—were selected. Our database thus consists of 100 images displaying three facial expressions by twenty actresses (no one actress was in more than one video). Ages of the actresses were not revealed in the (sales) texts describing the videos but they could be estimated via their facial features and body attributes; the actresses appeared to be within the 20 to 35-year age bracket.

2.2. Methods

We used AI image analysis software to: (a) extract a rectangle in each of the 100 frames containing the faces, plus negligible borders; (b) apply a suite of image analysis routines in the software package (MATHEMATICA® v12.4 from Wolfram Technology) to align the five faces for each actress in each video; (c) implement feature extraction algorithms to construct a feature vector for each face; (d) dimension-reduce the five feature vectors to five 2D vectors (which we call dimension-reduced feature vectors) for each actress's five facial expressions; and (e) calculate the Euclidean distances between pairs ("neutral"–"neutral", "neutral"–"pain", "neutral"–"pleasure", and "pleasure"–"pain") of these dimension-reduced feature vectors. Details and code of the software implementation are listed in Appendix A.

We consider the pain–pleasure dimension-reduced feature vector distances to have been drawn from a statistical population with an unknown parametric distribution. In this paper, we restricted ourselves to four parametric distributions: normal, log-normal, Gamma, and Weibull. We estimated the ML distribution [19] of the pain–pleasure distances and calculated both the mode and expectation value, as well as the highest density interval at 95% uncertainty ($HDI_{95\%}$) [20].

We attempted to quantify noise and looked for patterns as well as possible pattern structures after noise elimination. To do so, we generated a 20×5 matrix (in which the first three columns are the three distances of neutral from the neutral mean; the fourth column is the distance of pain from the neutral mean; and the last column is the distance of pleasure from the neutral mean). We scaled each column by its mean and centered it by subtracting a vector of 1s. We then performed a singular value decomposition (SVD) [21,22] to determine how many singular values contribute to the signal. We found that the sum of the first three singular values squared extracted 89.8% of the square of the Frobenius norm [21] of the matrix (which we call the scaled, shifted distance matrix). We then applied an unsupervised clustering algorithm (agglomeration cluster algorithm) to determine whether the rendition of affective states by the actresses are segmented into clusters.

3. Results

For each actress, the feature vector of each of the five frames was extracted using image analysis software. These five feature vectors for each actress were then dimension-reduced to 2D using an autoencoder (a neural network [22–24] with—in our case—seven layers; Figure 1). We also determined the 2D coordinates of the mean of the three neutral displays for each actress (because we argue that they have a commonality, although not necessarily equality). A result for one actress is shown in Figure 2.

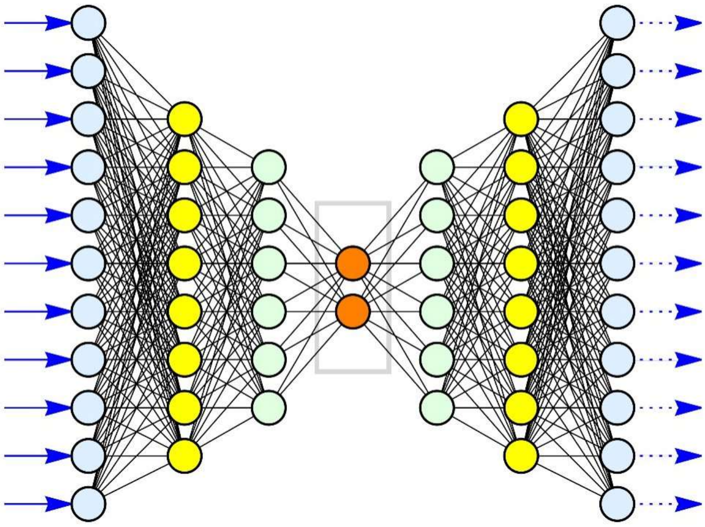


Figure 1. A symbolic rendition of an autoencoder used for dimension reduction. The 11 inputs (in the drawing) are represented by blue arrows from left to right. The inputs are conventionally labelled neurons (hence the name "neural network"). Each input neuron (light blue) has as many outputs as there are neurons in the next layer. Each 'yellow' neuron thus has 11 inputs/edges (represented as thin black lines). Each neuron in the yellow layer has as many outputs as there are neurons in the next layer: six outputs for each 'yellow' neuron and therefore eight inputs for each 'green' neuron of the next layer. It continues: each 'green' neuron has as many outputs as there are neurons in the next layer (consisting of two 'orange' neurons) and each 'orange' neuron has six inputs. The light blue neurons on the left are called the input layer while the light blue neurons on the right are called the output layer. The yellow layers, the green layers, and the orange layer are called the hidden layers. An autoencoder always has the same number of output neurons as it has input neurons. The number of hidden layers is part of the design by the engineer constructing the autoencoder. The numerical values along the black edges between neurons are determined by an algorithm. The autoencoder attempts to produce an output equal to the input (hence the name 'autoencoder') without being an identity mapping. An important feature for modern autoencoders is the ability to cut (set to zero) certain interconnections or make them numerically very small (usually by using a sigmoid function). The central layer is called the code. If the inputs are the feature vectors, then the numerical values of the code are the components of the dimension-reduced feature vector. If this is a successful autoencoder, it detects nonlinear combinations between the components of the (input) feature vector that can be represented by two variables.

We found that the pain–pleasure distances were rarely less than the pain–mean neutral and/or pleasure–mean neutral distances. In fact, the 2D points (mean neutral, pain, and pleasure) for most actresses often formed a near-isosceles triangle (occasionally one close to being equilateral).

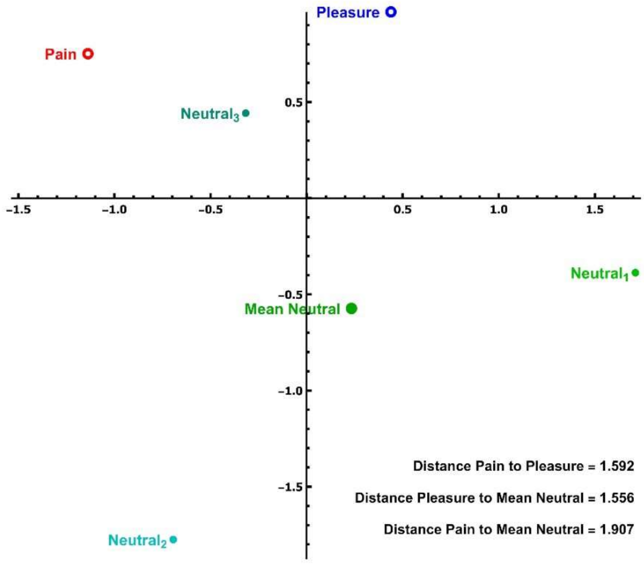


Figure 2. The locations of the dimension-reduced feature vectors of the facial display of the (labeled) affective states of one actress. In addition to these five points, we rendered (and used in subsequent calculations) the arithmetic mean (center of mass) of the neutral states, which we call 'mean neutral' for this female. Distances of the dimension-reduced feature vectors for this female are also displayed.

The distances between pain and pleasure were Weibull-distributed. The parameters of the ML distribution, its mode, expectation, and $HDI_{95\%}$ [20] uncertainty interval are listed in Table 1 and the pdf of this ML distribution is displayed in Figure 3.

Table 1. Parameters of the ML distribution of the pain–pleasure distances of the 20 actresses; it is a Weibull distribution. E is the expectation and $HDI_{95\%}$ is the 95% highest density interval [20].

Parameters	ML Numerical Values
WeibullDistribution$[k, \lambda]$	$k = 4.34$ $\lambda = 2.25$
Mode	2.12
E	2.05
$HDI_{95\%}$	$(s_1, s_2) = (0.99,\ 3.06)$

The sum of the first three singular values squared of the scaled, shifted distance matrix explains 89.8% of the square of the Frobenius norm of this matrix. We therefore looked for possible clusters in this 'smoothed' (noise-free) pattern matrix. We used the hierarchical agglomerative clustering algorithm and detected six clusters, of which the last two are isolates. Details of the clusters are described in the figure caption of Figure 4.

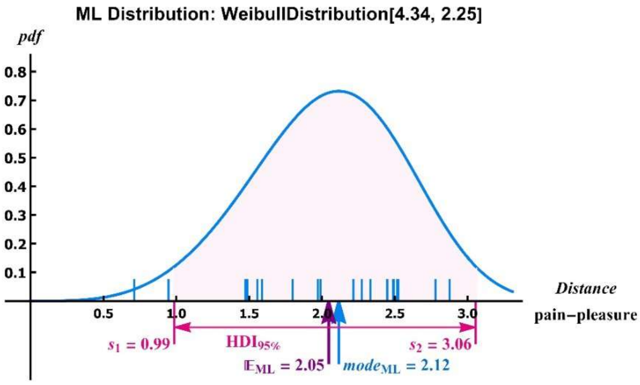


Figure 3. The ML distribution of the pain–pleasure distances for all 20 actresses. The distribution is asymmetric, thus displaying the interval mean ± SD is not meaningful. We used the $HD_{95\%}$ interval [20] to display the uncertainty. The ML mode and ML expectation are close to midway between the ends of the $HDI_{95\%}$ interval, and, furthermore, the ends of this interval are very close to $\frac{1}{2}\times$ and $\frac{3}{2}\times$ the mode. The shaded area under the pdf-curve has an area of 95%.

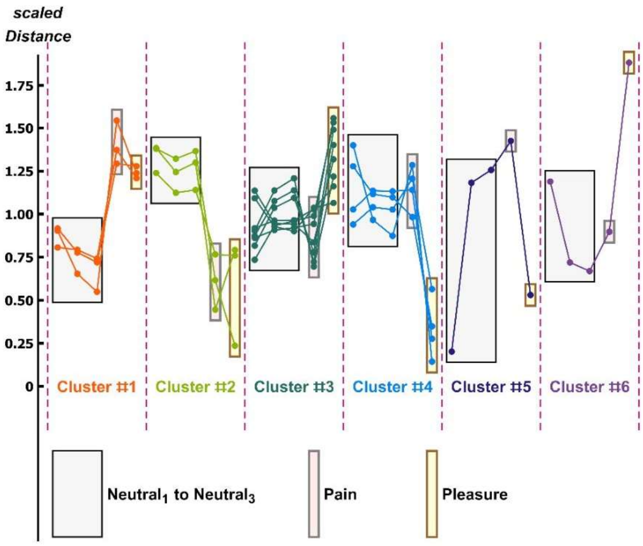


Figure 4. The clusters detected in the scaled, shifted pattern matrix. Distances are scaled by the neutral mean. There are six clusters with membership sizes: 3, 3, 8, and 4, along with two further clusters containing singletons (isolates). We observed the following: (i) In Cluster #1, the distances of pain and pleasure from the neutral mean are nearly equal and larger than those of the neutrals to the neutral mean. (ii) In Cluster #2, the distances of pain and pleasure to the neutral mean are also nearly equal but considerably less than the distances of the neutrals to the neutral mean. (iii) In Cluster #3, there is hardly any overlap between the pain and pleasure distances, and the pain distances overlap the neutral distances to the neutral mean. Additionally, the pleasure distances are larger than the pain distances. (iv) In Cluster #4, the pain distances overlap the neutral distances to the neutral mean, but the pleasure distances do not; the latter are, furthermore, much smaller than the other distances.

4. Discussion

The affective state distances between the dimension-reduced feature vectors of the expressions of pain and of pleasure are not small; they are, however, comparable with the distances between the three facial expressions of neutral. The facial expressions of pain and of pleasure are distinct for the image analysis software. We contrast this result with the observation that human raters cannot successfully distinguish between these facial expressions [15,25,26].

The three neutral faces of each actress are consistent within Cluster #1 and Cluster #2 but are inconsistent in Cluster #3 and Cluster #4, as well as inconsistent overall (Figure 4). After reviewing (by visual inspection) the frames used for the analysis, we noticed that the neutrality appeared to depend on the camera setting. This effect, which we could not rigorously quantify in this paper, is not a limitation. Raters of the actresses were also dependent, during their attempted identification of the facial expression of pain and pleasure, on the lighting and camera setting. Distances between the dimension-reduced feature vectors of each display of neutral are sometimes comparable to the distances of the pain–neutral mean or pleasure–neutral mean (Figure 4).

The affective state distances between the display of pain and the neutral mean varied considerably. Part of this variation, we discovered, was due to noise. Removal of noise via SVD resulted in distances that were members of one of four clusters (except for two isolates). It is remarkable, we argue, that the distributions of noise-free distances not only cluster but cluster into four (non-trivial) clusters. We suspect that the clustering is an effect ascribable not only to the camera setting but also to the commonalities of expression among the actresses in each cluster.

The human visual system (we discovered) is not well-equipped to rate the differences between the facial expressions of pain and pleasure, whereas the AI methods we present here are (as distances between the dimension-reduced feature vectors). We do note an added superiority of the AI image analysis, SVD, and hierarchical agglomerative clustering: we can quantify the noise and remove it prior to the identification of clusters. We do not argue that training removes errors. A recent publication [26] demonstrated that training (and expertise) does not remove all perception errors by humans.

Our analyses include a word of caution to researchers investigating the postulated indistinguishability of facial expressions of pain and of pleasure, especially if they are relying on data obtained via fieldwork, such as when evaluating ratings. The analyses presented here deals with how well image analysis software of facial expressions makes the distinction between the expression of pain versus that of pleasure possible. As we have documented, there is a small but not-to-be-neglected fraction of noise in the dataset. We must infer that this (statistical) noise is also responsible for making it more difficult for humans to correctly distinguish between these two affective states.

By implication, we conclude that our image analysis software's ability to distinguish between expressions of pain and of pleasure is superior to the human ability to do so.

5. Conclusions

We have discovered that image analysis software can be used to construct feature vectors of the facial expressions of pain and of pleasure, and compared them with the feature vectors of the neutral facial expressions.

Image analysis software of affective states has at least one helpful and one ominous use. The ominous one is unwanted supervision, which can be used to monitor the affective states of unsuspecting 'victims' (as in, for example, police states). The (numerous) helpful ones are those that allow for the monitoring of dangerous situations, such as distress, accident consequences, bodily harm, or situations of enforced, unwanted compliance (such as in violent situations). We repeat: properly developed image analysis software is needed to reliably distinguish between facial expressions of pleasure and pain.

The analyses of the outcomes provided considerable information. First, we found that AI methods can more reliably distinguish between pain and pleasure than humans can (thus succeeding in achieving our aim). Secondly, the reduced feature vector distances (the affective state distances) between pain and pleasure may be comparable to the pain to neutral distances, implying that identifying the facial expression(s) of affective states is difficult for humans. As the uncertainty interval of the pain–pleasure distance is very close to twice the mode of the pain–pleasure distance (Figure 3), we found a further explanation as to why it is so difficult for humans to visually distinguish between the facial expression of pleasure versus that of pain. Thirdly, the presence of noise contributes to the explanation of why humans have such difficulties when confronted with the task of distinguishing between facial expressions of pain and of pleasure. Fourthly, a clustering algorithm succeeded in identifying patterns in the noise-free pain–pleasure–neutral distances' renditions. Humans are not successful in seeing noise-free distances, thus we cannot expect humans to identify the discovered patterns in these distances.

Author Contributions: T.H. collected the videos and chose the frame sequences needed for subsequent analyses. J.B. and D.Ř. designed the research framework. H.P. designed the statistical tests and wrote the computer programs. H.P. and J.B. authored the manuscript. S.B. contributed to the research about human ratings of the affective states. All authors contributed to revisions after critically reading various versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study is a result of the research funded by the Czech Science Foundation, namely the project GA ČR 19-12885Y "Behavioral and Psycho-Physiological Response on Ambivalent Visual and Auditory Stimuli Presentation".

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Faculty of Science, Charles University, Prague, Czech Republic (protocol code 2018/08 approved on 2 April 2018).

Informed Consent Statement: Not applicable.

Data Availability Statement: The frames were extracted from commercially available videos. As the videos are proprietary, we can only make the extracted frames we used available (upon request) from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

We used routines supplied by Wolfram Research. We used MATHEMATICA v12.4 for this study.

In the (explicit) code below, the commands available in the MATHEMATICA package(s) are written in bold. Variables, parameters, etc., that are included by the authors in the code are not.

MATHEMATICA uses its own font, namely 'Consolas', which cannot be imported into this camera-ready document for all lines of the program code. The command lines are therefore written in Calibri and the explanatory text in Palatino Linotype. The explanatory comments are marked by bullet points, followed by a TAB.

- Every user has his/her own folder structure. By "*path*" (below), we mean the path to the folders containing the video frames.
- The command line **Join[. . .]** is long because it loads a segment of the video dynamically. It is modified accordingly for other videos loaded for further frame extraction.

SetDirectory["*path*"**]**
jpgApleasure **= Join[Flatten[Table[StringJoin[**"A_pleasure",**StringJoin[{ToString [0], ToString[i]}]**".jpg],{i,1,9}],**Flatten[Table[StringJoin[**"A_Pleasure",**StringJoin[{ToString[**j], **ToString[**i]}],{i,0,9}],{j,1,4}],{**StringJoin[**"A_Pleasure",**StringJoin[{ToString [5], ToString [0]}**],".jpg]}]
face **= Import[**jpgApleasure[[25]]]
faceApleasure **= FindFaces[**face,"Image",**Method**⟶"Haar",**PaddingSize**⟶30]

- The above structure is suitably modified for the other faces of Actress A.

faceAneutral = . . . ;
faceAneutral2 = . . . ;
faceAneutral3 = . . . ;
faceApain = . . . ;

- The five faces are aligned.

{faceAneutral,faceA,faceneutral2,faceApain,faceApleasure,faceAneutral3};
faceAjoin = **FaceAlign[**%,**Automatic**,{60,60},**PerformanceGoal**⟶"Quality"];

- The proprietary code from MATHEMATICA uses AI (internally trained) to extract feature vectors from the list of faces.

faceAextjoin **= FeatureExtraction[**faceAjoin];

- The proprietary code from MATHEMATICA uses a neural network to dimension-reduce the feature vectors.

faceAextractReduce = **DimensionReduce[**faceAextjoin,2,**RandomSeeding**⟶**Prime [137]**];
MeanAneutralReduce = **Mean[Table[**faceAextractReduce[[i]],{i,{1,2,5}}]];

- The (Euclidean) distances are computed.

distApain = **Norm[**faceAextractReduce[[3]]-meanAneutralReduce];
distApleasure = **Norm[**faceAextractReduce[[4]]-meanAneutralReduce];
distAPainPleasure = **Norm[**faceAextractReduce[[3]]-faceAextractReduce[[4]]];
neutralDistances = {**Table[Norm[**faceAextractReduce[[i]]-
Mean[Table[faceAextractReduce[[j]],{j,{1,2,5}}]]]],{i,{1,2,5}]
Table[Norm[faceTextractReduce[[i]]-
Mean[Table[faceTextractReduce[[j]],{j,{1,2,5}}]]]],{i,{1,2,5}]}

- The above steps are repeated for the other 19 × 5 faces; A⟶B, A⟶C, . . . and so on up to and including A⟶T.

painDistances = {distApain, ... distTpain}
pleasureDistances = {distApleasure, ... distTpleasure}
distancesPainPleasure = {distAPainPleasure, . . . , . . . ,distTPainPleasure}

- The commands below are used to find the ML distribution of the distances.

distributionList = {**NormalDistribution[**μ,σ], **LogNormalDistribution[**μ,σ], **WeibullDistribution[**k,λ], **GammaDistribution[**k,θ]};
data = distancesPainPleasure;
distNorm = **EstimatedDistribution[**data,distributionList[[1]]]
LLnorm = **LogLikelihood[**%,data]
distLogNorm = **EstimatedDistribution[**data,distributionList[[2]]]
LLlogn = **LogLikelihood[**%,data]
distWeib = **EstimatedDistribution[**data,distributionList[[3]]]
LLweib = **LogLikelihood[**%,data]
distGamm = **EstimatedDistribution[**data,distributionList[[4]]]
LLgamm = **LogLikelihood[**%,data]
logLikeList = { LLnorm,LLlogn,LLweib,LLgamm}
posML = **Flatten[Position[**%,Max[%]]][[1]]
distML = distributionList[[%]]

- The code below is used to determine the $HDI_{95\%}$ uncertainty interval. Note that the precision arithmetic requires several hundered (decimal) digits.

distA = %;
weib = distML;
modeWeib = $\left(1-\frac{1}{\%[[1]]}\right)^{\frac{1}{\%[[1]]}}\%[[2]]$
modeA = %;

$$\text{solutions3} = \ \mathbf{Reverse}[\mathbf{Table}[\mathbf{NSolve}[\{g == \mathbf{SetPrecision}[\text{i}(g/.u \rightarrow \text{modeA}), 5\mathbf{MachinePrecision}], 0 < u < 1\}, u, \mathbf{Reals}]//\mathbf{Quiet}, \{\text{i}, 0.004, 0.24, 0.002\}]];$$
$$\text{s1s2Solutions3} = \mathbf{Table}[\{u/.\text{solutions3}[[\text{i}, 2]], u/.\text{solutions3}[[\text{i}, 1]]\}, \{\text{i}, 1, \mathbf{Length}[\%]\}];$$
$$\text{cdfSolutions3} = \ \mathbf{Table}[\{\mathbf{CDF}[\mathbf{SetPrecision}[\text{distA}, 5\mathbf{MachinePrecision}], u/.\text{solutions3}[[\text{i}, 2]]] - \mathbf{CDF}[\mathbf{SetPrecision}[\text{distA}, 5\mathbf{MachinePrecision}], u/.\text{solutions3}[[\text{i}, 1]]]\}, \{\text{i}, 1, \mathbf{Length}[\%\%]\}];$$
$$\text{funcPts} = \mathbf{Transpose}[\{\text{cdfSolutions3}, \text{s1s2Solutions3}\}];$$
$$\text{iFun} = \mathbf{Interpolation}[\text{funcPts}, \mathbf{InterpolationOrder} \rightarrow 1];$$
$$\text{s1s2} = \mathbf{Sort}[\mathbf{SetPrecision}[\text{iFun}[0.95], 5\mathbf{MachinePrecision}]];$$

- The code below calculates the SVD and the approximation using only the first three singular values.

  matPre =
  Table[Flatten[{neutralDistances[[i]],painDistances[[i]],pleasureDistances[[i]]]],{i,1,20}**]**;
  Transpose[matPre];
  $$\text{matRed} = \mathbf{Transpose}[\mathbf{Table}[\frac{1}{\mathbf{Mean}[\%[[\text{i}]]]}\%[[\text{i}]] - 1, \{\text{i}, 1, 5\}]];$$
  $$\{\text{Umat3}, \text{Smat3}, \text{VTmat3}\} = \mathbf{SingularValueDecomposition}[\text{matRed}, 3];$$
  $$\text{mat3} = \text{Umat3.Smat3.}\mathbf{Transpose}[\text{VTmat3}];$$

- The code below generates a list of colors needed for the graphics.

  Delete[**ColorData**[3, ”ColorList”][[2; ;]], 2];
  Join[%,{**Darker**[Brown, 0.15]}, {Green}, {Cyan}**]**;
  Join[%,{**Darker**[Yellow, 0.15]}, {**Lighter**[Orange, 0.2]}, {Pink}];
  farbe2 = **Join[**%,{**Darker**[LightPurple, 0.1]}, {**Darker**[LightGreen, 0.35]}];

- The code below finds the clusters of the SVD-3 approximated coordinates of the affective state distances.

  clust = **FindClusters**[mat3, **Method** → ”Agglomerate”, **RandomSeeding** → **Prime** [137]];
  Table[**Length**[%[[i]]], {i, 1, Length[%]}]
  Length[%%]
  Flatten[**Table**[**Position**[mat3, clust[[j]][[i]]], {i, 1, Length[clust[[j]]]}]];
  clustPos = **Table**[**Union**[%[[j]]] , {j, 1, Length[clust]}];

- A suite of graphics routines (not listed) are used to display the results for the manuscript.

References

1. Prossinger, H.; Hladky, T.; Binter, J.; Boschetti, S.; Riha, D. Visual Analysis of Emotions Using AI Image-Processing Software: Possible Male/Female Differences between the Emotion Pairs “Neutral”–“Fear” and “Pleasure”–“Pain”. In Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference, Virtual Event, 29 June–2 July 2021; pp. 342–346.
2. Butow, P.; Hoque, E. Using artificial intelligence to analyse and teach communication in healthcare. *Breast* **2020**, *50*, 49–55. [CrossRef] [PubMed]
3. Hassan, T.; Seuß, D.; Wollenberg, J.; Weitz, K.; Kunz, M.; Lautenbacher, S.; Schmid, U. Automatic detection of pain from facial expressions: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1815–1831. [CrossRef] [PubMed]
4. Namba, S.; Sato, W.; Osumi, M.; Shimokawa, K. Assessing automated facial action unit detection systems for analyzing cross-domain facial expression databases. *Sensors* **2021**, *21*, 4222. [CrossRef] [PubMed]
5. Weitz, K.; Hassan, T.; Schmid, U.; Garbas, J.U. Deep-learned faces of pain and emotions: Elucidating the differences of facial expressions with the help of explainable AI methods. *Tm-Tech. Mess.* **2019**, *86*, 404–412. [CrossRef]
6. Dildine, T.C.; Atlas, L.Y. The need for diversity in research on facial expressions of pain. *Pain* **2019**, *160*, 1901. [PubMed]
7. Barrett, L.F. AI weighs in on debate about universal facial expressions. *Nature* **2021**, *589*, 202–204. [CrossRef] [PubMed]
8. Cowen, A.S.; Keltner, D.; Schroff, F.; Jou, B.; Adam, H.; Prasad, G. Sixteen facial expressions occur in similar contexts worldwide. *Nature* **2021**, *589*, 251–257. [CrossRef] [PubMed]
9. Ekman, P. Facial expression and emotion. *Am. Psychol.* **1993**, *48*, 384–392. [CrossRef]
10. Valente, D.; Theurel, A.; Gentaz, E. The role of visual experience in the production of emotional facial expressions by blind people: A review. *Psychon. Bull. Rev.* **2018**, *25*, 483–497. [CrossRef]

11. Gendron, M.; Barrett, L.F. Reconstructing the past: A century of ideas about emotion in psychology. *Emot. Rev.* **2009**, *1*, 316–339. [CrossRef] [PubMed]
12. van der Struijk, S.; Huang, H.H.; Mirzaei, M.S.; Nishida, T. FACSvatar: An Open Source Modular Framework for Real-Time FACS based Facial Animation. In Proceedings of the 18th International Conference on Intelligent Virtual Agents, Sydney, Australia, 5–8 November 2018; pp. 159–164.
13. Chen, C.; Crivelli, C.; Garrod, O.G.; Schyns, P.G.; Fernández-Dols, J.M.; Jack, R.E. Distinct facial expressions represent pain and pleasure across cultures. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E10013–E10021. [CrossRef]
14. Wenzler, S.; Levine, S.; van Dick, R.; Oertel-Knöchel, V.; Aviezer, H. Beyond pleasure and pain: Facial expression ambiguity in adults and children during intense situations. *Emotion* **2016**, *16*, 807. [CrossRef]
15. Aviezer, H.; Trope, Y.; Todorov, A. Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science* **2012**, *338*, 1225–1229. [CrossRef] [PubMed]
16. Fernández-Dols, J.M.; Carrera, P.; Crivelli, C. Facial behavior while experiencing sexual excitement. *J. Nonverbal Behav.* **2011**, *35*, 63–71. [CrossRef]
17. Hughes, S.M.; Nicholson, S.E. Sex differences in the assessment of pain versus sexual pleasure facial expressions. *J. Soc. Evol. Cult. Psychol.* **2008**, *2*, 289. [CrossRef]
18. Abramson, L.; Marom, I.; Petranker, R.; Aviezer, H. Is fear in your head? A comparison of instructed and real-life expressions of emotion in the face and body. *Emotion* **2017**, *17*, 557. [CrossRef] [PubMed]
19. Elisason, S.R. *Maximum Likelihood Estimation. Logic and Practice*; SAGE Publications: Newbury Park, CA, USA, 1993.
20. Kruschke, J.K. *Doing Bayesian Data Analysis. A Tutorial with R, JAGS, and Stan*, 2nd ed.; Academic Press/Elsevier: London, UK, 2015.
21. Leon, S.J. *Linear Algebra with Applications*, 5th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1998.
22. Strang, G. *Linear Algebra and Learning from Data*; Wellesley-Cambridge Press: Wellesley, MI, USA, 2019.
23. Russell, S.; Norvig, P. *Artificial Intelligence. A Modern Approach*, 3rd ed.; Pearson: Harlow, UK, 2010.
24. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*, 3rd ed.; MIT Press: Cambridge, MI, USA, 2016.
25. Boschetti, S.; Hladký, T.; Machová, K.; Říha, D.; Binter, J. Judgement of extreme affective state expressions in approach/avoidance paradigm. *Hum. Ethol.* **2021**, *36*, 7.
26. McKinney, S.M.; Sieniek, M.; Godbole, V.; Godwin, J.; Antropova, N.; Ashrafian, H.; Shetty, S. International evaluation of an AI system for breast cancer screening. *Nature* **2020**, *577*, 89–94. [CrossRef] [PubMed]

s

Vliv pornografie na přejímání emocí a sexuálního chování

Denisa Průšová[1] , Silvia Boschetti[2], Tomáš Hladký[1],Daniel Říha[1], Timothy Jason Wells[1] & Jakub Binter[1,2]

[1]Univerzita Karlova, Fakulta humanitních studií, Praha, Česká republika
[2]Univerzita Karlova, Přírodovědecká fakulta, Praha, Česká republika

Abstrakt

Emoce jsou důležitým aspektem v lidském vývoji i v sociálním chování a jejich správné čtení je kritické při interakci s ostatními, ať už se jedná o běžnou sociální interakci či romantické či sexuální chování. Kromě čtení emocí je důležitým aspektem také jejich přejímání, resp. zrcadlení na nevědomé úrovni, empatie Typicky je takovéto chování viditelné u dětí, kdy také dochází k vývinu těchto emotivních složek. Přejímání emocí však může být také negativní, jako je například duplikování agresivního chování. Co se týče sexuálního chování, může docházet k tomu, že toto chování je přejímáno z pornografických materiálů, kde se však jedná o emoce falešné, což vytváří mentální obraz určitého sexuální chování, který je považován za správný, avšak je zkreslen nároky (pornografické) filmové tvorby. Zejména důležitá je tato diskuze v oblasti sexuálního násilí, kdy zobrazované sexuální násilí či agrese v pornografickém materiálu může značně zkreslovat představy o povaze násilí. Tyto témata jsou diskutována zejména s ohledem na sexuální edukaci a vzdělávání.

Klíčová slova: *emoce, přejímání emocí, sexuální chování, pornografie, sexuální násilí*

Abstract

Emotions are an essential aspect of human psycho-social development, and their correct ascription is a crucial element when interacting with others. That goes for casual everyday interactions as well as romantic or sexual interactions. Reading emotions and emotional coupling or mirroring – empathy – is what makes such interaction understandable and pleasant. It is important, particularly in children who can absorb the emotional and behavioral responses to certain situations from adults, such as aggressive behavior observed in adults. Pornographic movies may be a source of imitation models. However, the emotional states depicted in the erotic material can be manipulated or even completely false, possibly creating a perception of appropriateness in a viewer. The images of violence can further grow in the imagination of the consumer. Therefore, the role of pornography should be discussed from a didactic point of view, which we attempt in this article.

Keywords: *emotion, emotional coupling, sexual behavior, pornography, sexual violence*

Úvod

Konzumování pornografie a sexuální násilí

Pornografie, tedy prezentace vzrušivých neuměleckých materiálů, které jsou vytvořeny za účelem vyvolání vzrušení u konzumenta, jejíž vkusnost a přípustnost je dána dobou, zákonem a obecným náhledem jednotlivce (kompilace několika definic upravená autory) je častěji sledována muži než ženami. Tento rozdíl se však v posledních letech snižuje (Ramlagun, 2012).

Sledování pornografických materiálů může ovlivňovat emoční odpověď na nevědomé bázi. Ačkoliv testované osoby nemusí reportovat subjektivní pociťování emoce při určité zobrazované emoci v pornografickém snímku, při laboratorním měření frekvenční odezvy mozku se ukazuje, že dochází k přejímání emocí podvědomě (Kunaharan et al., 2017).

Ačkoliv někteří jedinci mohou vykazovat sklony k sexuálnímu násilí při zvýšené konzumaci pornografie, není stále jednoznačné, zdali se jedná o kauzalitu. Výzkumy na tuto otázku nepředkládají jednoznačnou odpověď a často může být tato argumentace ovlivněna přístupem dané společnosti i konkrétního výzkumníka (např. Bachman et al., 1992; Bouffard, 2002). Nicméně problémem zůstává nedostatečné rozlišování mezi korelací a kauzalitou, což je bohužel stále častý jev.

Předešlé výzkumy dokonce naznačily, že zvýšené konzumování pornografie může vést k vyvolání negativních emocí, a tudíž i k sexuálnímu násilí (Wright et al., 2015; Paolucci et al., 1997). Sexuální predátoři přitom mohou konzumovat pornografii i signifikantně ve větší míře, než průměrný konzument (Johnson, 2015). I jiné studie ukazují užší souvislost mezi konzumováním pornografie, kde se jmenovitě vyskytuje nějaký násilný akt, a sexuálně motivovanými násilnými zločiny (Hald et al., 2010).

Jiné studie věnující se vztahu zvýšené konzumace pornografie a sexuálního násilí však nenašly žádnou signifikantní souvislost (např.: Allen et al., 1995; Ferguson & Hartley, 2009). Existuje také několik studií, které se spíše věnují negativním efektům zvýšené konzumace pornografie z hlediska emocí a škodlivého dopadu na sociální a potažmo sexuální chování. Zvýšená konzumace pornografie totiž může vést například ke zvýšeným úzkostem (Szymanski & Stewart-Richardson, 2014), depresím (Conner, 2014), či k potížím s erekcí při interakci s partnerem (Park et al., 2016).

Při častějším sledování pornografických materiálů také dochází k tomu, že konzumenti vyhledávají postupně čím dál tím víc explicitnější materiály, aby dosáhli stejné úrovně fyziologického vzrušení (Harper & Hodgins, 2016). Tedy jejich fyziologické vzrušení zůstává stejné, avšak je potřeba stále silnějších podnětů.

Jedná se o tzv. habituace, tedy navyknutí si na stejné prezentované stimuly, k čemuž následuje slabší sexuální odpověď; tento proces je přitom přítomný více u mužů než u žen (Koukounas & Over, 2000; Laan & Everaerd, 1995; O'Donohue & Plaud, 1991). Je možné, že tento proces je navíc přítomný i v reálných partnerských vztazích, kdy může docházet k navyknutí si na sexuální podněty, a tudíž i k případnému nesprávnému určování sexuálního vzrušení. Toto však může být obzvláště komplikované v případě, kdy u mužů dochází k sexuálnímu vzrušení a automaticky předpokládají stejný emoční stav u partnerky. Dochází tedy k nesprávnému čtení komunikovaných informací.

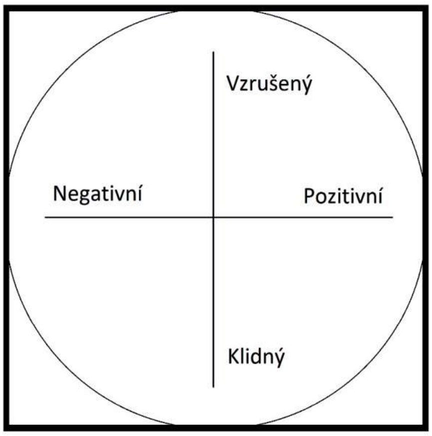
Určování emocí a dvoufaktorový model emocí

Schopnost odečítat výrazy tváře ostatních a dostávat tak tím informace o jejich momentálním mentálním stavu, náladě a emocích patří k základním a velmi důležitým prvkům v lidské komunikaci. Již Charles Darwin (1872) a později Paul Ekman (2006) upozorňovali na důležitost emocí a univerzálnost ve výrazu tváře. Odečítání emočních stavů z výrazu tváře se navíc neobjevuje pouze u člověka, ale i u jiných primátů člověku příbuzných (Darwin, 1872).

Určování emoce lze rozdělit do tří složek: a) *valence* (pozitivní a negativní, anglický originál: *Positive and Negative Affect*); b) *salience* (vysoká a nízká; značí míru prožívaného emočního stavu, značí míru vzrušení); c) kognitivní přiřazení (slouží k označení určitého psychofyziologického stavu jako „emoce"). Na základě kognitivního přiřazení spouštěče je následně psychofyziologický stav určen a označen jako emoce.

Emoce jako taková je nejenom prožívána, ale její prožívání je také některým z komunikačních kanálů vysíláno a přijímáno druhými osobami. Při takové komunikaci je však velmi důležitý vlastní emoční stav příjemce, který může významně ovlivňovat příjem a vnímání vysílané emoce. To se ukazuje zejména důležité při komunikaci s potenciálním partnerem či v již probíhajícím partnerském vztahu. Správné čtení a určení emoce mohou být kritické pro další interakci s potenciálním partnerem či k uchování současného vztahu. Emoce mohou být určeny i nesprávně.

V klíčových studiích Clarka a Sensibara z roku 1955 a Schachtera a Singera z roku 1962 bylo zjištěno, že při afektivních stavech může být k danému emočnímu stavu či vyladění přiřazen nesprávný význam. Konkrétně vzrušení z přívalu adrenalinu může být vyhodnoceno jako sexuální vzrušení. Autoři na základě těchto výzkumů vytvořili dvoufaktorovou teorii emocí (anglický originál: *Two-Factor Theory of Emotion*; Schachter & Singer, 1962) na jejímž základě byl později vytvořen Dvoufaktorový emoční model (Feldman, 1995; viz Obr. 1). Dle této teorie není fyziologická změna dostatečným podnětem, aby mohlo dojít k vnímání emoce. Kromě fyziologické změny je totiž také nutná přítomnost kognitivní složky. Tedy daný podnět v dané situace může zapříčinit, že je fyziologická odpověď přisuzována jinému podnětu, než který ji vyvolal.



Obr. 1: Dvoufaktorový emoční model Feldman, L. A. (1995)

Jedním z nejznámějších výzkumů, který se věnoval přejímání a kopírování emocí, jmenovitě tedy agrese a násilného chování, je experiment Alberta Bandury z roku 1961. Při něm jedna skupina dětí v předškolním věku sledovala videa, kde figurantka fyzicky napadala a nadávala hadrové panence. Této skupině pak byly ukázány hračky, se kterými si však děti nesmějí hrát. Pokud jim však byla představena panenka z videa, měly děti tendenci směřovat svou frustraci ze zákazu hraček a následnou agresivitu právě na onu panenku. Druhá skupina dětí, která videa neviděla, toto agresivní chování neměla. Došlo tedy k tomu, že děti replikovaly chování, emoce, které viděly, a které považovaly v tu chvíli jako za „správné", adekvátní. Zjednodušeně řečeno jako něco, co by se mělo dělat.

Na tomto místě také považujeme za důležité zmínit efekt tzv. emočního párování (*Emotional coupling*; Budell et al., 2010; Caruana, 2020), kdy mají jedinci tendenci přejímat komunikované emoce protějšku. Z hlediska sexuálního je to pak především sexuální vzrušení nebo naopak diskomfort, jejichž správné komunikování je důležité pro správné čtení situace (Aviezer et al., 2008; Clark & Sensibar, 1955; Younis et al., 2015).

Co se týče dlouhodobých partnerských vztahů, je toto emoční párování zejména důležité, jelikož je kritické pro spokojenost obou jednotlivců ve vztahu, tudíž i pro potenciální dlouhodobost tohoto vztahu. A právě sexuální složka partnerského vztahu je jedním z nejdůležitějších atributů (Montesi, 2011), přičemž muži obecně přikládají sexualitě ve vztahu větší důležitost než ženy (Peplau, 2003). Pro spokojenost ve vztahu je tudíž důležité správné určení původu sexuálního vzrušení, jelikož frekvence sexuální interakce je pro oba partnery stejná, ovšem prožívaná emoce či informace, které obě strany komunikují, mohou být odlišné (Birnbaum, 2006). To je dobře vidět na příkladu, že pokud mají ženy nižší svou vlastní sexuální vzrušivost, mají tendenci přisuzovat nižší spokojenost se sexuální interakcí ve vztahu. U mužů se tato tendence neobjevuje (Baumeister, 2000).

Pokud toto přejímání sexuální chování vztáhneme na sexuální chování u lidí a potažmo i na přejímání emocí, můžeme usoudit, že tento efekt může být velmi ovlivněn pornografickými materiály. Pornografie se stala běžnou součástí naší kultury (Paul, 2010; Weiss & Zvěřina, 2001) a pro dospívající jedince tak může představovat vlastně jediný způsob, jak přejímat sexuální chování. Co je však problematické je fakt, že sexuální chování a emoce zobrazované v těchto snímcích jsou většinou hrané, falešné. Scénáře jsou svou zápletkou nereálné. Zejména se pak jedná o nesouhlas se sexuálním stykem a záměrně zdůraznění překonání nesouhlasu. Scénáře totiž často představují vzrušující fantazie a touhy, které však nekorespondují s typickým sexuálním chováním. Svou charakteristikou jsou velmi extrémním zobrazením oproti typickému sexuálnímu chování. Sexualita je zde nadhodnocována a zdůrazňována. Emoce běžně vyskytující se v sexuální interakci v běžném životě či partnerském vztahu jsou tak většinou potlačovány nebo naopak přeháněny. Sexuální chování jednotlivců pak tímto může být velmi zkresleno a nekoresponduje s preferencemi a požadavky partnerů/partnerek.

Emoce v pornografii

Účinkování v pornografických snímcích je i pro herce a herečky významnou emoční prací (*Emotional labour*). Jde o aspekt pracovní činnosti, který odráží míru nutné emoční investice do zaměstnání, která je extenzí fyzického vykonávání zaměstnání (Parvez, 2006). Pro snadné pochopení bychom si tento pojem mohli vysvětlit na povolání servírky. Při práci není jejím jediným účelem zajistit objednávku, donesení jídla, kasírování a sklizení ze stolu. Její práce, navíc s ohledem na spropitné, spočívá také v příjemné komunikaci se zákazníkem, doporučení a nezřídka také v investici do vzhledu a vyjadřování (Erickson, 2004). Pravděpodobně všichni čtenáři si vybaví situaci, kdy byla obsluha nepříjemná a dojem z návštěvy zařízení tím trpěl. A také velmi dobře poznáme, že se obsluha přetvařuje. Stejně tak nekvalitní výkon u herců a o to víc pornoherců vybudí spíše výsměch nežli vzrušení. Proto je zvýšený tlak na to, aby byly snímky velmi realistické, a to zejména v extrémnějších polohách pornografe (jako např. BDSM snímky, tvrdý anální sex, dávení při orálním sexu, bolestné výkřiky při bičování, zavírání bradavek do svěráku, fackování a plácání atp.).

Pornografie krom toho totiž může působit jako tzv. supernormální stimul (*Supernatural stimuli*; Barrett, 2010). Supernormální stimul je koncept pozorovatelný v živočišné říši; využívá evolucí zafixované vnímání a preference a působí jako extrémně silný stimul díky svému charakteru. Typickým příkladem je kukaččí vejce. Kukačka snese do cizího hnízda své vejce, které je ovšem několikanásobně větší než vejce ptáka, který zde již své budoucí potomky nakladl. Daný druh ptáka máš však zafixovaný „stimul" vejce a mnohonásobně větší vejce na něj působí jako supernormální stimul, byť se jedná o vejce cizí. Díky svým extrémně odlišným proporcím vzbuzuje silný efekt a matka poté upřednostňuje toto vejce před svými vlastními a pečuje o něj více (Barrett, 2010).

Pornografie jako kompenzační chování

Je otázkou, zda by pornografie měla být tedy přístupná široké veřejnosti v takové formě jako doposud. Na základě výzkumu typů konzumentů pornografie se zdá, že takový přístup není problematický a že pornografie genericky neznamená problematický sexuální život, prožívání, nebo problémy ve vztazích. Je však zřejmé, že stejně jako tomu je u návykových látek jako alkohol, kde občasná uměřená konzumace může mít pozitivní účinky (ve formě mírného uvolnění a zvýšení prožitku), u určité populace může vyvolat závislost a může být nebezpečná jak konzumentovi, tak jeho okolí. Bylo by vhodné, a v jiných zemích již existují zaměřené organizace, vytvořit poradnu pro osoby, které mají s takovým chováním problém. V české republice podle osobních sdělení autorů od odborníků je takovéto chování řešeno v rámci sexuologické péče, pokud je potřeba, ale nejde o častý problém. V rámci preventivních programů je tomuto tématu věnován prostor, i když limitovaný.

Výsledky studie provedené vědeckým týmem vedeným Vaillancourt-Morel (2017) ukázaly, že konzumace pornografie je u jednotlivých skupin spojená s odlišnými riziky i benefity. Konkrétně rozdělila konzumenty na 3 rozdílné skupiny: asi 75 % bylo vyhodnoceno jako rekreační uživatelé, kteří spíše získávají inspiraci, a jejich sexuální život není negativně ovlivněn. Tato skupina je více zastoupena mezi ženami a zadanými jedinci obou pohlaví, s malým množstvím sexuálních problémů. Každý týden tato skupina průměrně 24 minut sledovala pornografické materiály. Skupina nekompulzivních uživatelů (13 %) i skupina kompulzivních uživatelů (12 %) byla více zastoupena muži. Skupina nekompulzivních uživatelů byla typická vysokým emočním stresem spojeným s konzumací pornografie, kdežto skupina kompulzivní měla středně silný emoční stres spojený s konzumací. Kompulzivní skupina také měla nejmenší poměr zastoupení zadaných jedinců.

Je však také nutné zvážit vliv extrémní pornografie jakožto kompenzačního chování. Ve studii na velkém vzorku (více než 5000 osob; Barker, 2014), byl diskutován vztah sexuálních fantasií a kompenzace reálné sexuální aktivity formou získávání v realitě potlačovaných prožitků, což dále přímo na případu pornografie ukazuje studie Efrati & Amichai-Hamburger z roku 2019 u studentské populace. Obě studie poukazují na mechanismy, které souvisí s prožíváním druhotného dopadu na vytváření sebe sama skrze sexuální chování bez interakce s jinou osobou, které jsou ovlivněny médií a konzumací materiálů s erotickou a pornografickou tématikou.

Závěr

Pornografie a její potenciální škodlivost či naopak prospěšnost vyvolává debaty i v současné době. Nejednoznačný efekt pornografie totiž vede k nezodpovězeným otázkám, jak sledování těchto materiálů může ovlivňovat sexuální chování. Především u dospívajících jedinců, kteří začínají svůj pohlavní život, může být jejich sexuální chování velmi ovlivněno právě chováním, které pozorují v pornografii, často negativním, resp. takovým, který je v nesouladu s nastavením a preferencemi protějšku. O to více je problematické, když jediným sexuálním chováním, které lze nezúčastněně pozorovat, je právě pornografie, která může vyvolávat zkreslenou představu o sexuálním životě. Tyto materiály jsou navíc často vytvořeny tak, že na lidské smysly působí jako supernormální stimul.

Debatu o sexuální výchově mladistvých by bylo vhodné vést spíše k edukaci o užívání a rizicích pornografie obecně. Pro rodiče/zákonného zástupce/dospělého/pro jedince samotného je těžké přesně vědět, co "jeho dítě" sleduje a "navíc vhodně posoudit", kde jsou hranice mezi běžným sexuálním chováním a extrémně zobrazovanými sexuálními praktikami. Je stále nutné omezování sledování a dostupnosti videí kvůli zachování nepřístupnosti snímků zobrazujících skutečné nelegální aktivity (např. dětská pornografie, skutečné extrémní nekonsensuální násilí).

Tato otázka je rovněž důležitá především v diskuzi o tzv. Rape myths, tedy stále se vyskytujícím nesprávném stereotypním předsudku o tom, že ženy si za znásilnění či jiné nevyžádané sexuální chování mohou samy. V pornografii je totiž velmi častým scénářem a tématem znásilnění. Pohlavní styk, který žena odmítne, je vynucen a ve většině případů žena nakonec na sexuální interakci přistoupí a vyjadřuje spokojenost. Zde je velmi problematická právě falešnost emocí. Jednak je zde velmi zkreslen zákonný koncept „ne znamená ne", kdy žena odmítá pohlavní styk pouze zdánlivě, a nakonec k němu přistoupí, a jednak je to nejednoznačnost emocí. Ta se týká jednak hraní emocí a jednak již zmíněného zdánlivého odmítnutí a následného přistoupení na sexuální akt a jeho prožívání.

Diskuze problematiky pornografie je dlouhodobějšího charakteru, avšak stále aktuální a velmi komplexní. Doufáme, že se nám v budoucích výzkumech podaří objasnit vztah čtení a komunikování emocí v pornografii, převážně sexuálního vzrušení a sexuální agrese a následné poznatky přispějí k celkové edukaci a povědomí o lidské sexualitě a vlivu nových technologií na ní.

Věříme, že zvýšení informovanosti o dopadu supernaturálních stimulů na prožívání sexuality jedince dovolí zvýšení míry benefitů ze sledování pornografie, a naopak ke snížení dopadů negativních.

Dedikace
Tento výstup je výsledkem badatelské činnosti podporované Grantovou agenturou České republiky v rámci grantu GA ČR 19-12885Y „Behaviorální a psycho-fyziologická reakce na prezentaci ambivalentních obrazových a zvukových stimulů".

Literatura

Allen, M., Emmers, T., Gebhardt, L., & Giery, M. A. (1995). Exposure to pornography and acceptance of rape myths. *Journal of Communication*, *45*(1), 5-26.
Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., ... & Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological science*, *19*(7), 724-732.
Bachman, R., Paternoster, R., & Ward, S. (1992). The rationality of sexual offending: Testing a deterrence/ rational choice conception of sexual assault. *Law and Society Review*, 343-372.
Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. The Journal of Abnormal and Social Psychology, 63(3), 575.
Barker, M. (2014). The 'problem'of sexual fantasies. *Porn Studies*, *1*(1-2), 143-160.
Barrett, D. (2010). *Supernormal stimuli: How primal urges overran their evolutionary purpose*. WW Norton & Company.
Baumeister, R. F. (2000). Gender differences in erotic plasticity: The female sex drive as socially flexible and responsive. *Psychological bulletin*, *126*(3), 347.
Birnbaum, G. E., Reis, H. T., Mikulincer, M., Gillath, O., & Orpaz, A. (2006). When sex is more than just sex: Attachment orientations, sexual experience, and relationship quality. *Journal of personality and social psychology*, *91*(5), 929.
Bouffard, J. A. (2002). The influence of emotion on rational decision making in sexual aggression. *Journal of Criminal Justice*, *30*(2), 121-134.
Budell, L., Jackson, P., & Rainville, P. (2010). Brain responses to facial expressions of pain: emotional or motor mirroring?. *Neuroimage*, *53*(1), 355-363.
Caruana, F. (2020). 3 Emotional Mirroring Promotes Social Bonding and Social Habits. *Habits: Pragmatist Approaches from Cognitive Science, Neuroscience, and Social Theory*, 79.
Clark, R. A., & Sensibar, M. R. (1955). The relationship between symbolic and manifest projections of sexuality with some incidental correlates. *The Journal of Abnormal and Social Psychology*, *50*(3), 327.
Conner, S. R. (2014). *Frequency of pornography use is indirectly associated with lower relationship confidence through depression symptoms and physical assault among Chinese young adults* (Doctoral dissertation, Kansas State University).
Darwin, C. (1872). *The expression of the emotions in man and animals*, NY: D. Appleton and Company
Efrati, Y., & Amichai-Hamburger, Y. (2019). The use of online pornography as compensation for loneliness and lack of social ties among Israeli adolescents. *Psychological reports*, *122*(5), 1865-1882.
Ekman, P. (2006). Darwin and facial expression: A century of research in review. NY: Academic Press.
Erickson, K. (2004). To invest or detach? Coping strategies and workplace culture in service work. *Symbolic Interaction*, *27*(4), 549-572.
Feldman, L. A. (1995). Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of personality and social psychology*, 69(1), 153.
Ferguson, C. J., & Hartley, R. D. (2009). The pleasure is momentary… the expense damnable?: The influence of pornography on rape and sexual assault. *Aggression and violent behavior*, *14*(5), 323-329.
Hald, G. M., Malamuth, N. M., & Yuen, C. (2010). Pornography and attitudes supporting violence against women: Revisiting the relationship in nonexperimental studies. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *36*(1), 14-20.
Harper, C., & Hodgins, D. C. (2016). Examining correlates of problematic internet pornography use among university students. *Journal of Behavioral Addictions*, *5*(2), 179-191.
Johnson, S. A. (2015). The role of pornography in sexual offenses: Information for law enforcement & forensic psychologists. *Int. J. Emerg. Ment. Health Hum. Resil*, 17, 239-242.

Koukounas, E., & Over, R. (2000). Changes in the magnitude of the eyeblink startle response during habituation of sexual arousal. *Behaviour research and therapy*, *38*(6), 573-584.

Kunaharan, S., Halpin, S., Sitharthan, T., Bosshard, S., & Walla, P. (2017). Conscious and non-conscious measures of emotion: Do they vary with frequency of pornography use?. *Applied Sciences*, 7(5), 493.

Laan, E., & Everaerd, W. (1995). Determinants of female sexual arousal: Psychophysiological theory and data. *Annual Review of Sex Research*, *6*(1), 32-76.

Montesi, J. L., Fauber, R. L., Gordon, E. A., & Heimberg, R. G. (2011). The specific importance of communicating about sex to couples' sexual and overall relationship satisfaction. *Journal of Social and Personal Relationships*, *28*(5), 591-609.

O'Donohue, W., & Plaud, J. J. (1991). The long-term habituation of sexual arousal in the human male. *Journal of Behavior Therapy and Experimental Psychiatry*, *22*(2), 87-96.

Paolucci, E. O., Genuis, M., & Violato, C. (1997). A meta-analysis of the published research on the effects of pornography. *Medicine, Mind and Adolescence*, *72*(1-2).

Park, B. Y., Wilson, G., Berger, J., Christman, M., Reina, B., Bishop, F., ... & Doan, A. P. (2016). Is Internet pornography causing sexual dysfunctions? A review with clinical reports. *Behavioral Sciences*, *6*(3), 17.

Paul, P. (2010). From pornography to porno to porn: How porn became the norm. *The social costs of pornography: A collection of papers*, 3-20.

Parvez, Z. F. (2006). The labor of pleasure: How perceptions of emotional labor impact women's enjoyment of pornography. Gender & Society, 20(5), 605-631.

Peplau, L. A. (2003). Human sexuality: How do men and women differ?. *Current directions in psychological science*, *12*(2), 37-40.

Ramlagun, P. (2012). "Don't call me weird, but I normally watch porn"-Girls, sexuality and porn. *Agenda*, *26*(3), 31-37.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological review*, *69*(5), 379.

Szymanski, D. M., & Stewart-Richardson, D. N. (2014). Psychological, relational, and sexual correlates of pornography use on young adult heterosexual men in romantic relationships. *The Journal of Men's Studies*, *22*(1), 64-82.

Vaillancourt-Morel, M. P., Blais-Lecours, S., Labadie, C., Bergeron, S., Sabourin, S., & Godbout, N. (2017). Profiles of cyberpornography use and sexual well-being in adults. *The journal of sexual medicine*, *14*(1), 78-85.

Weiss, P., Zvěřina, J. (2001). *Sexuální chování v ČR – situace a trendy*. Portál, Praha.

Wright, P. J., Tokunaga, R. S., & Kraus, A. (2016). A meta-analysis of pornography consumption and actual acts of sexual aggression in general population studies. *Journal of Communication*, *66*(1), 183-205.

Younis, I., El-Esawy, F., & Abdel-Mohsen, R. (2015). Is female orgasm an earth-moving experience: an Egyptian experience. *Human Andrology*, *5*(3), 37-44.

Kontakt:
binter@email.cz

Bolest a slast v sexuálním chování

Jakub Binter[1,2], Silvia Boschetti[2], Tomáš Hladký[1], Daniel Říha[1] & Timothy Jason Wells[1]

[1]Univerzita Karlova, Fakulta humanitních studií, Praha, Česká republika
[2]Univerzita Karlova, Přírodovědecká fakulta, Praha, Česká republika

Abstrakt

Z evolučního hlediska jsou bolest a slast jedním z nejdůležitějších procesů pro přežívání živočišných druhů. Indikace bolesti napovídá organismu, že se má aktu, který mu bolest způsobuje, vyhnout, že pro něj představuje nebezpečí a že je pro něj škodlivá. Naopak slast přináší organismu uspokojení a říká mu, co je pro něj přínosné a příjemné. Slast tak může přinášet téměř cokoliv, jako je například uspokojení z dobrého jídla, které je rovněž důležité pro přežití. Avšak ne vždy jsou tyto dva procesy vnímané jako rozdílné a protichůdné. Jsou oblasti, kdy jeden z prožitků může způsobit či ovlivňovat druhý. Konkrétně co se týká oblasti lidské sexuality, je mnohdy rozdíl mezi těmito dvěma velmi intenzivními stavy stírán; prolínají se a vydávají se, obrazně řečeno, na stejnou cestu. Bolest totiž může přinášet, nebo potencovat i slast. To má za následek vyhledávání těchto velmi intenzivních stavů.

Kľúčové slová: *bolest, slast, sexuální vzrušení, lidská sexualita*

Abstract

From evolutionary point of view, pain and pleasure are some of the most important processes for survival. The pain serves as indicator that the causing activity or stimulus should be avoided for its potential to cause harm. Conversely, pleasure is pleasing and shall guide the organism or individual towards repetition and benefit. There are various stimuli that may bring pleasure such as food that is also beneficial for survival. Thus the two processes are not always distinct and countering each other. There are moments when one experience can cause or potentiate the other. Concretely, human sexuality is a field where the two can coexist and share the paths. Pain may bring or potentiate gratification and pleasure. This may be the final reason why people intentionally search for these intensive experiences.

Keywords: *pain, pleasure, sexual arousal, human sexuality*

Úvod

Evoluční pohled

Vnímání bolesti je komplexní fenomén, který má v centrální nervové soustavě významnou odpověď. Dá se říci, že v lidském mozku spolu neustále zápasí, ovšem se i navzájem doplňují, prožitky bolesti a prožitky slasti. Jsou to totiž dvě motivace, jelikož vyhledávání slasti či něčeho, co přináší příjemné pocity a vyhýbání se bolesti a nepříjemným pocitům, jsou jedny z klíčových motivací pro přežití. Bolest a slast tedy bývají v odborných výzkumech zkoumány jako dva rozdílné intenzivní stavy, ovšem propast mezi nimi není až tak veliká, jak se může zdát (např. Seymour a kol., 2007; Scott a kol., 2007; Fields, 2007).

To že je bolest významným činitelem v lidském chování si můžeme snadno demonstrovat na případu osob, které trpí mutací, která jim znemožňuje cítit bolest. Tato mutace dokonce významně ovlivňuje jejich prožívání strachu a stresu – které se sekundárně vyvíjí na základě prožité bolesti (Habib a kol., 2019).

Bolest je navíc velmi dobře studované téma, jako mnohé negativní emoce u člověka. Je jednoduché (fyzickou bolest) vyvolat v laboratorních podmínkách, je relativně univerzální, co se stimulů týče (Edwards a kol., 2004), i přesto, že existují rozdíly v závislosti na osobnostních charakteristikách (Diener a kol., 2009). Strach a hněv vyvolává tendence utéct a bojovat (Tooby a Cosmides, 1990; McCarty, 2016). Naproti tomu existuje široké spektrum pozitivních stimulů, které jsou velmi kulturně ovlivněné, a existuje velmi málo stimulů, které jsou pozitivní a naprosto univerzální. Jako příklad si můžeme uvést vazbu na jídlo, které obsahuje vysoký obsah tuků a cukrů, které byly v lidské evoluční historii klíčové pro přežití (Tooby a Cosmides, 1990; Drewnowski a kol., 1992). Vyvolávání slasti, je v laboratorních podmínkách velmi těžké s výjimkou podávání psychoaktivních látek (Haney, 2009) a sexuálního vzrušení (Hamilton a kol., 2008). Naše porozumění těmto mechanismům je ovšem velmi malé v porovnání s prožitky negativními, a to ani nezmiňujeme bolest a slast psychickou způsobenou prožitkem.

Bolest, tedy fyzický diskomfort může být i jakýmsi mezistupněm ke slasti, resp. k potěšení ve formě odměny. Dobrým příkladem je forma bolesti u sportovců, kteří jsou schopni, za vidinou odměny vítězstvím, tuto bolest částečně potlačit (Bentham, 1996). Slast je tak logickým vyústěním předešlého utrpení, které však zpětně jako takové utrpení vnímané není a je „přebito" pozitivním prožitkem z vítězství. Prožitek slasti může být rovněž tím větší, čím větší je předešlá bolest. Typickým příkladem je obranný mechanismus u organismů – čím je hrozba pro organismus větší, tím větší bývá i bolest, která agresivněji spustí obranné mechanismy a mechanismy snažící se hrozbě vyhnout (Price a kol., 1987). Tento stav se nazývá homeostáze, jedná se o rovnováhu vnitřních procesů v těle organismu. Jakmile dojde k vybočení z této rovnováhy negativním směrem (bolest), snaží se organismus tuto rovnováhu vyrovnat opačným pocitem, tedy slastí (Craig, 2003).

V tomto případě lze hovořit také o tzv. Motivation-Decision Model (model motivace a rozhodnutí), který funguje na bázi nevědomých rozhodovacích procesů (Fields, 2006). Do tohoto modelu pak spadají informace o vnitřním stavu organismu, tedy o homeostatické rovnováze, jako je pocit hladu, senzorické vjemy a povědomí o přítomné hrozbě a potenciální odměně. Dle Motivation-Decision Model tak teoreticky cokoliv, co je důležitější pro přežití, má potlačit způsobovanou bolest a pozornost má být směřována převážně na přežití.

Bolest a lidská sexualita – sadomasochistické chování

Konkrétní oblastí, ve které figuruje bolest a slast, a které se chceme věnovat, je lidská sexualita. Prožitek z bolesti totiž může být natolik intenzivní, že přináší sexuální vzrušení, potažmo je schopen přivodit orgasmus ve chvíli jejího náhlého odeznění (Mayberry a Daniel, 2016). Vztah sexuálního vzrušení a bolesti či diskomfortu lze nejlépe zkoumat na sadomasochistickém chování, tedy sexuálním chování, které je založeno na přijímání či způsobování bolesti, na cíleném ponižování či na dominanci a submisivitě.

V úvahu by připadaly ještě studie o sexuálním násilí, ovšem nejedná se samozřejmě o konsenzuální styk a k sexuálnímu vzrušení nedochází u obou stran. Některé zdroje navíc uvádějí, že agrese působí naopak jako „anti-vzrušivý" element, a tudíž není kompatibilní se sexuálním vzrušením, jelikož neuropřenašeč nazývaný sympatický katecholamin, který souvisí právě s hněvem a agresí, rovněž zapříčiňuje ústup sexuálního vzrušení u mužů, penilní detumescenci (Hnasko a kol., 2005; Pecina a kol., 2003). Ve studiích, které se zabývaly sexuálními zločiny, v některých případech k erekci vůbec nedocházelo (Barbano a Cador, 2006; Barbano a Cador, 2007). Vzrušení ze sexuálního násilí – mluvíme-li o nekonsenzuálním styku – nicméně častý jev je. Zde ho ovšem rozebírat nebudeme.

Ačkoliv je sadomasochistické chování považováno některými zdroji za sexuální poruchu, v tomto textu to tak vnímáno nebude a bude přijímán názor SM komunit, tedy že se nejedná o poruchu a případné deviantní chování, ale o sexuální preferenci (Wismeijer a van Assen, 2013). Rovněž z hlediska psychoanalýzy se sadomasochismus považuje za psychopatologii, ovšem sociologické a sociálně psychologické výzkumy takový postoj nezastávají – jedná se o osoby emočně a psychicky vyrovnané, které přijaly svou sexuální roli a nevybočují z rolí sociálních (Weinberg, 2006). Jedním z nejdůležitějších aspektů je však konsenzus. Není možné provozovat sexuální sadomasochistické praktiky bez souhlasu všech zúčastněných (Lawrence a Love-Crowell, 2008; Lindemann, 2011; Kolmes, Stock a Moser, 2006).

Je však nutné poznamenat, že ne vždy je při sadomasochistických praktikách přítomna bolest. Pod pojem sadomasochismus mohou spadat rozličné praktiky – pro tyto praktiky je často používán pojem BDSM, který se skládá ze tří zkratek – B/D (bondage and discipline neboli svazování a poslušnost/výcvik), D/s (dominance a submisivita) SM (sadismus a masochismus). (Kolmes, Stock a Moser, 2006).

Rovněž způsobování či přijímání bolesti nemusí prvoplánově sloužit právě za účelem bolesti. Účelem může být vyjádření hierarchického rozdělení, tedy k vyjádření dominantní a submisivní role (Cross a Matheson, 2006; Wismeijer a van Assen, 2013). Jak tedy lze vidět, v těchto případech nemusí sexuální uspokojení či slast pocházet ze samotné bolesti či způsobování bolesti, ale právě z rozdílného postavení účastníků sexuálních praktik. Jedinci praktikující tuto formu erotických her tyto aktivity mohou vykonávat spíše kvůli smyslovému uspokojení, a ne tolik kvůli erotickému uspokojení (Newmahr, 2010).

Jak již bylo zmíněno výše, slast může být způsobena úlevou od bolesti. Toto by bylo možné aplikovat i na sadomasochistické chování. Slast ani tak nepřináší samotná bolest, jako právě uvolnění od bolesti. Čím je bolest větší, tím větší je úleva, když bolest pomine a tím větší je i vzrušení z tohoto zážitku. Stále se však jedná o intenzivní stavy, kdy jeden doplňuje druhý, ani v tomto případě nejsou protichůdné.

Sadomasochismus je v dnešní době přijímán téměř jako normativní sexuální chování (Weinberg, 2006; Wismeijer a van Assen, 2013), kdy je brán a provozován bez různých předsudků, kterým čelil v minulosti a dostává se jak do běžného života, tak do mainstreamových médií (Wilkinson, 2009). Sexuální vzrušení z bolesti, popřípadě z dominance a ponižování (případně přijímání bolesti a submisivity) je totiž jev, ke kterému má každý člověk blízko. Bolest, poslušnost a oddanost je brána jako „zpestření" „klasického" sexuálního styku a paradoxně je to právě bolest, která je nejjednodušeji „dostupná". Ze všech lidských prožitků má nejblíže ke vzrušení a je v podstatě velmi jednoduché ji provozovat. Existuje několik vysvětlení, proč je tímto prožitkem právě bolest.

Nejednoznačné prožitky – nesprávné určení vzrušení

Do schématu vzrušení z bolesti může také spadat nesprávné určení příčin vzrušení (anglický originál: *Misattribution of Arousal*), kdy nesprávné čtení emocí a jejich následný přenos může ovlivňovat vnímání celé situace.

Například v klíčové studii Schachtera a Singera z roku 1962 měli mužští účastníci výzkumu za úkol přejít jeden ze dvou mostů společně s ženskou výzkumnicí. Jeden z mostů byl zavěšen nízko, druhý vysoko a houpal se. Následně vyplňovali tzv. tematický apercepční test (přisuzování významů nejednoznačným obrázkům) a měli možnost ve dnech následujících po testu kontaktovat výzkumnici pro případ jakýchkoliv dotazů. Ukázalo se, že pokud muži přecházeli po vysoko zavěšeném mostě, měli tendenci častěji přisuzovat obrázkům sexuální význam a častěji zpětně kontaktovali spolupracovnici. Pokud tedy došlo k nespecifickému vzrušení (v podobě adrenalinu z přecházení mostu), měli participanti tendenci vykládat toto vzrušení jako sexuální a přikládali mu sexuální význam.

Je tedy možné, že tento fenomén je přítomný i při sadomasochistickém chování, respektive v takovém chování, kdy je vzrušení z bolesti připsáno sexuálnímu vzrušení. Je však zřejmé, že tento fenomén nenastává u všech jedinců. Někteří jedinci mohou být na toto nespecifické vzrušení citlivější, a tudíž právě jim pak může bolest přinášet slast. Kromě výše popsaného fenoménu však do celé problematiky spadá i lidský mozek.

Bolest a lidský mozek

Z výzkumů totiž vyplývá, že požitek bolesti i požitek slasti v naší nervové soustavě aktivují insulární kortex, tedy oba požitky aktivují stejnou část mozku (Casey a spol., 2001; Komisaruk a spol., 2004). S použitím magnetické rezonance bylo dokonce zjištěno, že při orgasmu je v mozku aktivních až 30 oblastí, a to včetně těch, které jsou aktivní i při bolesti (Komisaruk a kol., 2011). Může tedy docházet k interakci těchto dvou požitků, pokud jsou v této části mozku aktivovány, což by vysvětlovalo podobnost výrazu tváře bolesti a slasti – viz níže (Komisaruk a spol., 2006).

Kontext je tak často velmi důležitým prvkem, dle kterého lze tyto dvě emoce od sebe rozdělit. Nicméně v určitých případech si tyto dva pocity konkurují nebo přinejmenším mohou nastat ve velmi krátkých intervalech po sobě. Vrátíme-li se k sexuálně motivovanému násilí, mnoho sexuálních násilníků si často vyloží emoční stav oběti jako konsenzuální a sexuálně žádoucí, nikoliv jako stav strachu a bolesti (Johnson, 2006).

Bolest a slast v lidské tváři

Z evolučního hlediska je rovněž zajímavé, že výrazy bolesti a slasti u člověka se velmi podobají a často je složité je od sebe rozeznat, ačkoliv, jak již bylo řečeno výše, jsou tyto dva velmi intenzivní stavy kritické pro přežití organismu, a to právě i co se týče výrazu tváře druhých.

Odečítání emočních stavů z výrazů tváře ostatních se objevuje i u ostatních primátů, kteří jsou člověku příbuzní (Darwin, 1872).

Výrazy bolesti mohou varovat ostatní před hrozbou nebo naopak vyvolávají empatické pocity, a tudíž zvyšují šanci na přežití ve formě pomoci od ostatních (Williams, 2002). Fyzické vyjádření požitku slasti je tak klíčem k úspěšnému rozmnožení, potažmo k uchovávání vztahu.

Dalším zajímavým příkladem je zjištění, že pozitivní výrazy tváře jsou rychleji rozpoznány než ty negativní, a to obzvláštně výrazy ženské tváře (Grimshaw a spol., 2004; Hugenberg a Sczesny, 2006). Ovšem toto platí spíše při rozpoznávání individuálních tváří. Pokud je například více tváří v davu, jsou zde rychleji rozpoznány výrazy agrese, tedy negativní emoce, a více přitahují pozornost (Eastwood a kol., 2001). Toto je zřejmě určené tím, že rozpoznat výraz agrese ve velkém davu zvyšuje šanci na varování před potenciální hrozbou a na včasné reagování (Hansen a Hansen, 1988).

Ženy jsou přitom zdatnější ve správném odečítání emocí z výrazů tváře ostatních (Hall a Matsumoto, 2004; Hampson, van Anders a Mullin, 2006), ovšem muži lépe a rychleji rozpoznají z výrazu tváře hněv (Rotter a Rotter, 1988). Nicméně, studie Chena a kolegů z roku 2016 zjistila, že výrazy tváře bolesti a slasti naopak od sebe odlišitelné jsou. Respondenti správně určili emoci, což by značilo, že ačkoliv objektivně tyto dvě emoce je těžké rozlišit, podvědomě poznáme jejich správný význam z kontextu.

Výrazy tváře, ať již před započetím sexuálních aktivit, či v jejich průběhu, jsou klíčové pro prožívání sexuálního vzrušení nebo pro vytváření vazeb mezi partnery a měly by být studovány s ohledem na přesnost určení. V BDSM komunitách je pro případ nesprávného pochopení prožitků druhé osoby určeno gesto (například zdvihnutí ruky), nebo tzv. safe-word (slovo, při jehož vyřčení je jakákoli další aktivita zastavena), což dovoluje stupňování negativních prožitků, které jsou součástí těchto aktivit (Dancer a kol., 2006). Alternativou může být tzv. semafor, kdy partner, který přijímá submisivní pozici, může vyřčením barvy (zelená = mám zájem pokračovat intenzivněji, oranžová = zůstaňme nejvýše v této intenzitě, červená = okamžitě sniž intenzitu / zastav) intenzitu prožitku regulovat. Je však otázkou, nakolik je takový přístup praktikován v párech, které nemají předem nastavené hranice a je nutné odečítat emoce pouze z tváře či hlasového projevu.

Závěr

Z uvedeného vyplývá, že je velmi důležité vnímat stavy bolesti a slasti nikoliv jako protichůdné prožitky; nebo respektive chápat jejich vzájemné propojení. Nelze je však ani vnímat, že jeden prožitek spouští druhý, tedy že se nutně střídají. Tyto dva velmi intenzivní stavy totiž spíše koexistují spolu, vyvolávají podobné reakce a navzájem se doplňují. To je z velké části podmíněno tím, že bolest i slast vyvolávají velmi podobné fyziologické reakce těla, o čemž svědčí i aktivování stejných oblastí mozku. Fyziologická reakce těla na vzrušení je velmi podobná reakci na stresové podněty a organismu, i nám subjektivně, mohou tyto dva „požitky" připadat stejné; „slévají" se dohromady a je těžké subjektivně, objektivně, i za pomoci laboratorních přístrojů určit původ výsledného prožitku. Je potřeba dalšího výzkumu pro pochopení vzájemné interakce těchto prožitků, což ztěžuje fakt, že některé observační metody v laboratorních podmínkách nedovolují správné odečtení emoce z výrazů tváře. Naturalistický výzkum a kvalitativní sběr dat ve spolupráci s komunitami, které se těchto výzkumů účastní, je tak klíčový pro budoucí pokrok v této oblasti a předem jim za jejich ochotu děkujeme.

Dedikace

Tento výstup je výsledkem badatelské činnosti podporované Grantovou agenturou České republiky v rámci grantu GA ČR 19-12885Y „Behaviorální a psycho-fyziologická reakce na prezentaci ambivalentních obrazových a zvukových stimulů".

Literatura

Barbano, M. F., & Cador, M. (2006). Differential regulation of the consummatory, motivational and anticipatory aspects of feeding behavior by dopaminergic and opioidergic drugs. *Neuropsychopharmacology*, *31*(7), 1371-1381.

Barbano, M. F., & Cador, M. (2007). Opioids for hedonic experience and dopamine to get ready for it. *Psychopharmacology*, *191*(3), 497-506.

Bentham, J. (1996). *The collected works of Jeremy Bentham: An introduction to the principles of morals and legislation*. Clarendon Press.

Casey, K. L., Morrow, T. J., Lorenz, J., & Minoshima, S. (2001). Temporal and spatial dynamics of human forebrain activity during heat pain: analysis by positron emission tomography. *Journal of Neurophysiology*, *85*(2), 951-959.

Craig, A. D. (2003). A new view of pain as a homeostatic emotion. *Trends in neurosciences*, *26*(6), 303-307.

Cross, P. A., Matheson, K. (2006). Understanding Sadomasochism. Journal of Homosexuality, 50:2, 133-166.

Dancer, P. L., Kleinplatz, P. J., Moser, C. (2006). 24/7 SM slavery. Journal of Homosexuality, 50, 81-101.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London: J. Murray.

Diener, E., Sandvik, E., & Pavot, W. (2009). Happiness is the frequency, not the intensity, of positive versus negative affect. In *Assessing well-being* (pp. 213-231). Springer, Dordrecht.

Drewnowski, A., Krahn, D. D., Demitrack, M. A., Nairn, K., & Gosnell, B. A. (1992). Taste responses and preferences for sweet high-fat foods: evidence for opioid involvement.

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception & psychophysics*, *63*(6), 1004-1013.

Edwards, R. R., Haythornthwaite, J. A., Sullivan, M. J., & Fillingim, R. B. (2004). Catastrophizing as a mediator of sex differences in pain: differential effects for daily pain versus laboratory-induced pain. *Pain*, *111*(3), 335-341.

Fields, H. L. (2007). Understanding how opioids contribute to reward and analgesia. *Regional Anesthesia & Pain Medicine*, *32*(3), 242-246.

Grimshaw, G. M., Bulman-Fleming, M. B., & Ngo, C. (2004). A signal-detection analysis of sex differences in the perception of emotional faces. *Brain and cognition*, *54*(3), 248-250.

Habib, A. M., Okorokov, A. L., Hill, M. N., Bras, J. T., Lee, M. C., Li, S., ... & Ramirez, J. D. (2019). Microdeletion in a FAAH pseudogene identified in a patient with high anandamide concentrations and pain insensitivity. *British journal of anaesthesia*, *123*(2), e249-e253.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*(2), 201.

Hamilton, L. D., Rellini, A. H., & Meston, C. M. (2008). Cortisol, sexual arousal, and affect in response to sexual stimuli. *The journal of sexual medicine*, *5*(9), 2111-2118.

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behavior*, *27*(6), 401-416.

Haney, M. (2009). Self‐administration of cocaine, cannabis and heroin in the human laboratory: benefits and pitfalls. *Addiction biology*, *14*(1), 9-21.

Hansen, C. & Hansen, R. D. (1988). Finding the face in a crown: An anger superiority effect. *Journal of Personality and Social Psychology*, 54, 917-924.

Hnasko, T. S., Sotak, B. N., & Palmiter, R. D. (2005). Morphine reward in dopamine-deficient mice. *Nature*, *438*(7069), 854-857.

Hugenberg, K., & Sczesny, S. (2006). On wonderful women and seeing smiles: Social categorization moderates the happy face response latency advantage. *Social Cognition*, *24*(5), 516-539.

Chen, C., Crivelli, C., Garrod, O., Fernandez-Dols, J. M., Schyns, P., & Jack, R. (2016). Facial expressions of pain and pleasure are highly distinct. *Journal of Vision*, *16*(12), 210.

Johnson, S. A. (2006). *Physical abusers and sexual offenders: Forensic and clinical strategies*. CRC Press.

Kolmes, K., Stock, W., Moser, Ch. (2006). Investigating bias in psychotherapy with BDSM clients. Journal of Homosexuality, 50, 301–324.

Komisaruk, B. R., Beyer-Flores, C., & Whipple, B. (2006). *The science of orgasm*. JHU Press.

Komisaruk, B. R., Whipple, B., Crawford, A., Grimes, S., Liu, W. C., Kalnin, A., & Mosier, K. (2004). Brain activation during vaginocervical self-stimulation and orgasm in women with complete spinal cord injury: fMRI evidence of mediation by the vagus nerves. *Brain research*, *1024*(1-2), 77-88.

Komisaruk, B. R., Wise, N., Frangos, E., Liu, W. C., Allen, K., & Brody, S. (2011). Women's clitoris, vagina, and cervix mapped on the sensory cortex: fMRI evidence. *The journal of sexual medicine*, *8*(10), 2822-2830.

Lawrence, A. A., Love-Crowell, J. (2008). Psychotherapists' Experience with Clients Who Engage in Consensual Sadomasochism: A Qualitative Study. Journal of Sex & Marital Therapy, 34, 67–85.

Lindemann, D. (2011). BDSM as therapy? Sexualities, 14 (2), 151–172.

Mayberry, L., & Daniel, J. (2016). 'Birthgasm' A Literary Review of Orgasm as an Alternative Mode of Pain Relief in Childbirth. *Journal of Holistic Nursing*, *34*(4), 331-342.

McCarty, R. (2016). The fight-or-flight response: A cornerstone of stress research. In *Stress: Concepts, cognition, emotion, and behavior* (pp. 33-37). Academic Press.

Newmahr, S. (2010). Rethinking kink: Sadomasochism as serious leisure. Qual Sociol, 33, 313–331.

Pecina, S., Cagniard, B., Berridge, K. C., Aldridge, J. W., & Zhuang, X. (2003). Hyperdopaminergic mutant mice have higher "wanting" but not "liking" for sweet rewards. *Journal of Neuroscience*, *23*(28), 9395-9402.

Price, D. D., Harkins, S. W., & Baker, C. (1987). Sensory-affective relationships among different types of clinical and experimental pain. *Pain*, *28*(3), 297-307.

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior*, *12*(2), 139-148.

Scott, D. J., Stohler, C. S., Egnatuk, C. M., Wang, H., Koeppe, R. A., & Zubieta, J. K. (2007). Individual differences in reward responding explain placebo-induced expectations and effects. *Neuron*, *55*(2), 325-336.

Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *Journal of Neuroscience*, *27*(18), 4826-4831.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. Psychological review, 69(5), 379.

Tooby, J., & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and sociobiology*, *11*(4-5), 375-424.

Weinberg, T. S. (2006). Sadomasochism and the Social Sciences: A Review of the Sociological and Social Psychological Literature, Journal of Homosexuality, 50 (2), 17-40.

Wilkinson, E. (2009). Perverting visual pleasure: Representing sadomasochism. *Sexualities*, *12*(2), 181-198.

Williams, A. C. D. C. (2002). Facial expression of pain: an evolutionary account. *Behavioral and brain sciences*, *25*(4), 439-455.

Wismeijer, A. A. J., van Assen, M. A. L. M. (2013). Department. Psychological Characteristics of BDSM Practitioners 1943. J Sex Med, 10, 1943-1952.

Konzumace pornografie obyvatel ČR a její vztah k demografickým proměnným: klastrování pomocí umělé inteligence

Hermann Prossinger[1], Silvia Boschetti[2, 3], Tomáš Hladký[2, 3] & Jakub Binter[2, 3]

[1]Oddělení evoluční antropologie, Fakulta přírodních věd, Vídeňská univerzita, Vídeň, Rakousko
[2]Fakulta humanitních studií, Univerzita Karlova, Praha, ČR
[3]Přírodovědecká fakulta, Univerzita Karlova, Praha, ČR

Abstrakt

Současný stav poznání: Ve společnostech s vyspělou ekonomikou a prakticky neomezeným přístupem k internetu (jako je Česká republika) je anonymní konzumace pornografie (často i obsahující prvky sado-masochismu) běžnou součástí života. Není však známo, jaké části populace se vystavují jakému typu těchto materiálů a zda taková aktivita souvisí s jejich vzděláním, frekvencí masturbace a do jaké míry je závislá na pohlaví.

Tato analýza dosud ani nebyla možná, s ohledem na dostupnou výpočetní sílu a na ní závislé postupy statistického zpracování dat. Díky užití umělé inteligence můžeme tyto prvky analyzovat všechny najednou a skutečně tak zjišťovat souvislosti mezi nimi.

Cíle: Zjistit souvislosti mezi různými proměnnými (pohlaví, partnerský status, vzdělání) a nepartnerským sexuálním chováním (konzumaci běžných i sado – masochistických pornografických materiálů a frekvenci masturbace). Očekávali jsme, že se účastníci výzkumu rozdělí do jednotlivých segmentů s jasně definovanými vlastnostmi. Následně jsme tyto segmenty porovnali s demografickými údaji, konkrétně velikostí místa bydliště.

Výsledky: Pomocí naší analýzy jsme zjistili, že participant se rozdělí do 7 jasně definovaných segmentů. Žádný ze segmentů neobsahuje pouze jedince jednoho pohlaví, většina z nich však obsahuje velkou převahu jednoho pohlaví. Frekvence autosexuálních praktik se významně liší v jednotlivých segmentech, stejně jako konzumace pornografických materiálů obou typů. Vzdělání, oproti našemu očekávání, není významným rozdělujícím prvkem mezi segmenty, stejně tak partnerský status. Vysoká konzumace pornografie, společně s vysokou frekvencí masturbace, charakterizuje dva segmenty.

Abstract

Background: In societies with advanced economies and ubiquitous internet access (such as the Czech Republic), anonymous consumption of porno videos (both BDSM and non-BDSM) has become widespread. However, it is neither clear what fraction of consumers in such a society (and its demographic stratification) consume how much, and how this relates to their education level, masturbation frequencies and to what extent it is sex-dependent.

Aim: First, we are interested in associations between various personal attributes (sex, relationship status, education) and sexual attributes (BDSM and non-BDSM porn video consumption as well as masturbation frequencies). We expect the participants to cluster into distinct segments with well-identifiable features. Second, we suspect that the segments of personal and sexual attributes associate with a demographic parameter, namely the size of the village, town, city in which they live.

The analytical tools were by now unavailable due to computational power and availability of neural-network driven Artificial Intellingence based analytical tools.

Results: We find that the participants cluster into 7 segments; none are exclusively of only one sex, but several are overwhelmingly of one sex only. The spectrum of masturbation frequencies is distinctly different in the segments, as are consumption of porno videos of both types. Education levels and relationship statuses are not good separators. Highly compulsive porno video consumption, together with high masturbation frequencies, characterize two segments, while restraining (conservative) sexual attitudes are characteristic of two others. Associations with demography are as expected: members with conservative sexual attitudes are primarily female and live in small towns and villages, as do males with highly impulsive sexual

Další dva segmenty jsou charakterizovány spíše konzervativním přístupem k sexualitě.

Závěr: Nové a pokročilé formy statistické analýzy dat dovolují identifikaci klastrů dospělých konzumentů pornografie na základě jejich auto-sexuálního chování a demografických proměnných, které by bez užití neuronových sítí nebylo možné identifikovat.–

Kľúčové slová: *Neuronové sítě – algoritmus DBSCAN – konzumace pornografie – frekvence masturbace – demografické parametry – úroveň vzdělání – kompulzivní sexuální chování.*

attitudes.

Conclusion: Modern, advanced statistical methods can identify segments of adults with well-defined sexual attributes and associations with demography that would otherwise remain undetected..

Keywords: *Neural networks – DBSCAN algorithm – porno video consumption – masturbation frequency – demography parameters – education status – compulsive sexual behavior.*

Úvod

"Zaručeně největší a nejsilnější ze všeho zla na světě je to, které vyrůstá z kořenů lidské přirozenosti. Není tomu tak snad, že tato část našeho bytí, jenž se pevně váže k základním silám přírody a jí formované přirozenosti člověka, ta se ukázala být nejvíce zvrácená – neukazuje to snad na veliký problém?" (Hopkinsová, 1899; překlad Binter, 2021).

Puritánství, jehož byla Ellice Hopkinsová zásadní představitelkou poukázalo na budoucí trend v lidské sexualitě. Z některých mužů, kteří jsou daleko snadněji vizuálně stimulovaní než ženy, se stali věrní konzumenti (příležitostně se hovoří i o kompulzivní konzumaci až nelátkové závislosti) internetové pornografie. Avšak platí za to náležitou cenu – ovlivněním sexuálního vývoje, stejně jako preferencí a sexuálního chování. U některých jedinců hrozí riziko přetrvávání narušení psychosexuálního vývoje i v dospělosti. Deset procent britského vzorku dokonce tvrdilo, že je přímo závislá na konzumaci pornografie (Logue et al., 2018).

Zvláště u mužů dochází k velkému překrývání sexuálních fantazií a preferencí skutečně sledované pornografie (Chivers & Bailey, 2005; Chivers et al., 2010; Binter et al., 2013; Krejčová et al., 2018). To zřejmě souvisí s dalšími dvěma aspekty: (a) vyhledávání sexuálního zážitku zobrazeného v takových materiálech a s tím související sexuální touhy (Galther & Sellborn, 2003; van Lankveld et al., 2015), a (b) dostupnost sexuálního partnera, se kterým by měl takové sexuální zkušenosti (Binter et al., 2012).

Dostupnost sexuálního partnera, který by splňoval požadavky konzumentů specifických typů pornografických materiálů, se dá spíše předpokládat ve větších městech, kde budou také pravděpodobněji bydlet jednotlivci s vyšším vzděláním (jelikož vysoké školy jsou převážně ve větších městech; Binter et al., 2012). Je rovněž pravděpodobné, že ve větších městech se vyskytuje více osob s menšinovými preferencemi (jako jsou BDSM-pozitivní jedinci; Richters et al., 2008; Wismeijer & van Assen, 2013). Důvody jsou pravděpodobně anonymita, větší partnerský trh, více klubů a lokalit věnovaných okrajové sexualitě.

Dalším vlivem na kvalitu partnerství a uspokojení sexuální touhy, která je často spojována s množstvím konzumované pornografie, je délka partnerského vztahu. Jedinci, kteří jsou v partnerském vztahu déle, s větší pravděpodobností fantazírují o jiných sexuálních partnerech než o těch, se kterými jsou právě ve vztahu. Dále pak jedinci, kteří podváděli své předchozí partnery, a ti, kteří dříve měli více sexuálních partnerů, častěji uplatňují toto své chování ve svých současných vztazích (Binter et al., 2012, Hicks & Leitenberg, 2001). Naopak ti partneři, kteří fantazírují o svých partnerech, budou pravděpodobně více projevovat chování související s partnerskou spokojeností – jak ukazují předchozí studie (Birnbaum et al., 2019). Nevyřešenou otázkou je však to, zda nižší spokojenost ve vztahu souvisí právě s tendencí k podvádění a sledování pornografie či jde o dva spíše nezávislé fenomény.

Důvodů ke sledování online pornografie je několik, například sexuální vzrušení; inspiruje také zvědavost z vlastního uspokojení z rozličných pornografických scén, a zvědavost ohledně technik uspokojování partnera (Sabina et al., 2008). Data z České republiky ukazují rozdíl mezi těmi, kteří jsou ve vztahu, a těmi, kteří jsou nezadaní (Smutny & Sulc, 2018). Nezadaní jedinci častěji (17,1 % vs. 12,5 %) sledují pornografii, aby uspokojili své tajné touhy; hledání inspirace pro sexuální život je však nižší o 13,2 % (21 % nezadaných uvádí takovou motivaci). Obě skupiny jsou motivovány ke sledování pornografie sexuálním uspokojením (v partnerství: 59,9 %; nezadaní: 66,7 %).

Může se zdát, že pornografie má pouze negativní dopady. Existují také pozitivní vlivy - kompenzační účinek, který byl prokázán v Japonsku, Dánsku, Finsku, Chorvatsku, Spojených státech amerických a České republice (Diamond et al., 2011). Zvýšená dostupnost pornografických materiálů je v dané zemi negativně spojena se zločiny souvisejícími se sexuálním násilím. Jinými slovy tam, kde je pornografie dostupná se sexuálně motivovaná kriminalita významně snížila.

Kromě toho se zdá, že pornografie je mysl otevírajícím faktorem, a to zejména pro ženy, které v minulosti nesledovaly pornografii, a bylo málo pravděpodobné, že ji někdy sledovat budou. Nyní však pornografii konzumují. Pornografie ženám pomáhá přesněji definovat, jaké jsou jejich sexuální potřeby, kontextualizovat tyto potřeby a preference a snáze přijmout vlastní sexualitu (McKeown et al., 2018).

Dalším vlivem, který je třeba zohlednit je to, jak často jedinec pornografii konzumuje. V českém vzorku, který je nejrelevantnější pro srovnání s tím, který v této studii prezentujeme my, byla pornografie konzumována velmi zřídka malým procentem mužů (3,4 %), ale přibližně jednou třetinou (29,5 %) žen. Více než polovina žen konzumuje pornografické materiály několikrát za měsíc, zatímco více než polovina mužů konzumuje pornografii 2 až 5krát týdně. Méně než jedno procento žen konzumuje pornografii alespoň jednou denně. 11 % mužů konzumuje pornografické materiály několikrát denně a 14,4 % alespoň jednou denně (Smutny & Sulc, 2018).

Stručná historie vizuální pornografie

Vyobrazování pornografie bylo zdokumentováno: (a) v peripatetickém Řecku, (b) během Římské říše (stále existující pornografické výjevy lze vidět nejen v nevěstinci v Pompejích, ale také na freskách ve vilách, které byly z velké části zničeny erupcí Vesuvu v roce 79 n. l. a nyní jsou vystaveny v Archeologickém muzeu v Neapoli), (c) v čínské kultuře (například Zhou Fang, 8. století) a v japonských (shunga) dřevorytech, (d) v indické Kamasutře (Sinha, 1988), jakož jako (e) ve známých sochách na fasádách indických chrámů (například chrám Kandariya Mahadeva, 11. století).

V čase poskočíme dále, abychom mohli porovnat možnosti přístupu k pornografickým mteriálům v období renesance v Itálii 16. století a biedermeierovského období (po Vídeňském kongresu 1815) v habsburské monarchii. Oba tyto případy nám ukazují snížení exkluzivity tohoto artiklu a dále pak zvýšení a usnadnění distribuce.

Zatímco Raphael (1483–1520) maloval své světoznámé fresky ve Stampe ve Vatikánu, jeden z jeho asistentů Romano (1499–1546; který se měl sám stát velmi slavným malířem) namaloval na zdi řadu lascivních kopulačních poloh. Z nich Marcantonio Raimondi (1482–1534) vyrobil měděné lepty a distribuoval je v Římě. To rozzuřilo Papeže a uvrhl Raimonda do vězení. Nicméně tato zobrazení byla poté znovu reprodukována a distribuována v knižní podobě, doprovázená extrémně explicitními pornografickými sonety napsanými Pietrem Aretinem (1492–1556).

Tyto I Modi (nejlépe lze přeložit jako kopulační pohyby, které zahrnovaly vyobrazení a popisy análního styku – sodomie nebyla ve vyšších vrstvách společnosti vždy kriminalizována) byly opakovaně skrývány a ničeny, když byly nalezeny, a podobně. V roce 1928 objevil Walter Toscanini (syn slavného dirigenta) exemplář z počátku 16. století s dřevoryty místo měděných leptů, který byl posléze prodán jinému soukromému sběrateli. O 62 let později byly publikovány reprodukce (Lawner, 1984). V kontextu tohoto článku je důležité poznamenat, že lepty a dřevoryty kolovaly pouze mezi bohatými a mocnými v Římě a Benátkách, a byly často přijímány společností.

Během Vídeňského kongresu (1815) bylo erotické (a pornografické) umění distribuováno v tajnosti (a zajisté vášnivě přijímáno). Do obecného povědomí se dostala série 40 akvarelů zobrazujících kopulační scény, které byly falešně připisovány Peteru Fendimu (1796–1842). Na rozdíl od I Modi v době renesance byly tyto materiály mnohem snáze dostupné jak pro prohlížení, tak i k pořízení do vlastnictví a obchodování, a to i lidmi z nižší šlechty a vrstvy zámožných prostých občanů. Protagonisty kopulačních scén byli také Vídeňané střední třídy oblečeni v biedermeierovském stylu (i když jen částečně kvůli pornografickým vyobrazením) – což je značný kontrast k oblečení šlechty té doby.

V Evropě a Severní Americe se pornografie po osvícenství (od poloviny 19. století) stávala stále více všudypřítomným a masově konzumovaným artiklem. S vynálezem fotografie a hlavně filmu, byla umožněma prezentace vizuální pornografie přístupná stále větším částem populace – především díky inherentním technikám masové produkce a masové distribuce filmových materiálů. (První společnost, která výhradně produkovala erotické a pornografické filmy, byl Saturn-Film, Vídeň, v letech 1907–1911).

V dnešní době (2021) vedlo zaplavení pornografickými filmy doprovázené jejich snadnou anonymní distribucí ke zdánlivě nekritické a téměř neomezené spotřebě. V tomto článku však ukazujeme, že spotřeba není tak nekritická, jak se běžně věří (viz níže).

Vývoj BDSM (bondage-dominance-sadismus-masochismus) chování a jeho vizuálních prvků

BDSM je žánr a životní styl. Musíme však rozlišovat mezi aktivitami BDSM, které jsou čistě sexuální, a těmi, které zahrnují sexualitu, avšak jsou jakýmsi souhrnem sociálních interakcí (Foucault, 1990). Raným vyobrazením bičování se sexuálním podtextem je Tomba della Fustigazione (Hrobka bičování), etruská hrobka nacházející se v Itálii, datovaná kolem roku 490 př. n. l., zobrazující jednotlivce, kteří prožívají (sexuální) slast a zároveň jsou bičováni a plácáni rukou (spanking) a bičem (Steingräber, 2006). Poté následuje velká časová mezera v hmotných důkazech tohoto vyobrazování. Jedno z přežívajících ztvárnění je Satyr bičující nymfy uvázané ke stromu, rytina Antonia Carracciho (1557–1602). Donatien Alphonse François, markýz de Sade, napsal 120 dní Sodomy a Justinu; oba romány považované za klenoty BDSM literatury, ve vězení. Většina literárních popisů je však pravděpodobně čistě smyšlená, ale ze soudních řízení a soudních obvinění víme o faktických popisech jeho preferencí, za které byl stíhán zejména ve svém mládí (Gray, 1998).

Richard Freiherr von Krafft-Ebing byl první, kdo ve svém díle Psychopathia Sexualis (1886) nazval takové chování – působení jinému bolesti za účelem vlastního potěšení – jako sadismus a přidal i další pojem, masochismus. Masochismus je pojmenován po Leopoldovi von Sacher-Masochovi, rakouském šlechtici, který strávil mládí v Praze. Jeho spisy, stejně jako jeho sexuální život, se zaměřovaly na záměrně přijímanou bolest, tresty a degradaci. Jeho Venuše v kožichu (1870) zahrnuje pasáže oddanosti, sexuálního násilí, psychického týrání a fetišismu.

Moderní klinické chápání sadismu a masochismu je považováno za nekonsensuální a liší se od Sacher-Masochova. Konsenzuální sadomasochismus je naproti tomu v poslední době považován za variantu sexuální preference bez negativních důsledků (Federoff, 2008).

Moderní vizuální styl a charakter konsensuálního BDSM lze hledat u jediného fotografa ze čtyřicátých a padesátých let minulého století, který vytvořil vzhled, jenž s tématem BDSM od té doby spojujeme. Byl jím Johna Alexandra Scotta Cuttsa (pseudonym John Willie; Pine, 2013).

V současné době pornografická videa s prvky BDSM zobrazují sexuální chování, během kterého dochází k (obvykle konsensuální) výměně moci. Některé praktiky jsou zaměřeny na fyzické vjemy, některé na prožitky a některé na duševní stavy (včetně například trestu a hraní rolí). Bolest přitom může, ale nemusí být přítomna. Myšlenkou je spíše poslušnost a podřízenost. Někdy se k zintenzivnění prožitku používají speciální kostýmy (Binter et al., 2021).

Věk nových technologií

Jeden z největších a život měnících technologických pokroků, vznik internetu, má na svědomí obrovské množství pornografických materiálů. Nabízí tak pestrou paletu materiálů, že snad žádné pornografické téma nechybí. Každé páté internetové vyhledávání je zaměřeno na pornografii; každý čtvrtý respondent výzkumu má ve svém mobilním zařízení uložené pornografické materiály; jedna třetina obsahu internetu celkově je pornografická (Smutny & Sulc, 2018). Přibližně 70 % spotřebitelů používá ke sledování pornografie notebook a nejméně jedna třetina jej sleduje na chytrém telefonu. 90 % obou pohlaví vždy sleduje něco nového, zatímco méně než 25 % mužů a 10 % žen sleduje pornografické materiály vícekrát. Většina ne-konzumentů pornografie měla za celý život jednoho nebo dva sexuální partnery; téměř 1/5 z konzumentů měla více než deset. To dále ukazuje na celkové propojení sexuality soukromé a sdílené s partnerem / partnery.

Muži, kteří jsou více zaměřeni na vizuální prvky, častěji sledují pornografii téměř všech typů a témat (včetně, ale ne pouze, materiály zaměřené na fetiš, skupinový sex, sexuální interakce dvou žen). U dvou typů pornografických témat v jednom z výzkumů však ženy převyšují muže: (a) v BDSM pornografii, kde dominuje muž a (b) v pornografii, kde je přítomné zobrazení předehry (softcore pornografie). Nebyly nalezeny žádné rozdíly v pohlaví pro zobrazení „běžných" (někdy nazývaných „obvyklých" nebo „konvenčních") sexuálních aktivit (Krejčová et al., 2018).

Jak jsme poznamenávali výše, je účelem sledování pornografie nejčastěji spojované s masturbací. Reprezentativní vzorek české populace ukazuje, že u obou pohlaví je nástup masturbace obvykle ve věku 13 let a frekvence se zvyšuje do 15 let, kdy již tuto aktivitu provozuje pravidelně většina populace (Weiss & Zvěřina, 2001). Předpokládá se, že frekvence masturbace má negativní dopad na spokojenost se sexuálním životem (Brody & Costa, 2009), ale odborníci nejsou v tomto jednotní; v kontextu konzumace pornografie a masturbace probíhá dále diskuze o jejím dopadu na sexuální chování jak v pozitivním, tak negativním ohledu (Dines, 2010; Gyamfi & Zalata, 2018, Boschetti et al., 2021; Průšová et al., 2021).

V kontinuální studii sexuálního chování v České republice (Weiss & Zvěřina, 1993; 1998; 2003; 2008) bylo zjištěno, že muži masturbují častěji než ženy (muži 7,1/měsíc, ženy 5,8/měsíc). V roce 1993 potvrdilo pravidelnou masturbaci 84 % mužů, 83 % v roce 1998, 89 % v roce 2003 a 90 % v roce 2008. 50 % žen potvrdilo masturbaci v roce 1993, 58 % v roce 1998, 61 % v roce 2003 a 65 % v roce 2008. Trvalý nárůst frekvence je u obou pohlaví podobný. 76 % mužů považuje masturbaci za přirozený jev; stejně tak 62 % žen. 20 % mužů považuje masturbaci za špatný, i když neškodný, zvyk, stejně jako 29 % žen. 4 % mužů a 9 % žen považuje masturbaci za škodlivou.

Výše uvedené charakteristiky nám dokládají existenci rozličných tříd konzumentů pornografických materiálů. Vaillancourt-Morel (2017) například identifikoval tři: (a) rekreační uživatele (přibližně 75 %), (b) kompulzivní uživatele (13 %) a (c) ne-kompulzivní uživatele, ne-rekreační uživatele (12 %). Uživatelé rekreační (a) uvedli, že jejich hlavní motivací je vlastní inspirování a že jejich sexuální život nebyl negativně ovlivněn konzumací porna. Ženy byly převládajícími členy této skupiny a strávily přibližně půl hodiny týdně sledováním pornografie. Okolo 13% kompulzivních uživatelů však chápe svou konzumaci jako problematickou.

V další studii bylo zjištěno, že 10,2 % ze vzorku vysokoškolských studentů a 7,8 % kontrolního vzorku je kompulzivními uživateli. Šlo obvykle o heterosexuální muže (především mladší jedince) s klinickými příznaky kompulzivní poruchy sexuálního chování (Castro-Calvo et al., 2020).

Cíle a metody studie

Tento článek se zabývá různým (ale ne nerozlišujícím) stupněm spotřeby pornografie námi přijaté dichotomie (BDSM pornografie versus pornografie bez prvků BDSM) a tím, jak tato konzumace souvisí s dalšími parametry ve dvou oblastech: a) demografické kategorie (velikosti místa bydliště) ve kterých respondenti žijí) a b) osobní charakteristiky (rodinný stav, úroveň vzdělání a rovněž frekvence masturbace).

Naše metoda analýzy, využívající metody umělé inteligence a klastrovací algoritmus bez učitele (unsupervised clustering algorithm) – vzorem není rozhodování člověka, ale vlastní kategorizace na základě dostupných informací. Výsledky následně budeme porovnávat s výše publikovanými výsledky dosud publikovaných relevantních studií.

Materiály

Výzkum týkající se sexuálních aktivit probíhal online a celkem se ho zúčastnilo 902 respondentů z České republiky. Z toho dotazník vyplnilo celkem 881 osob (Tab. 1).

Charakteristika	Kód	Popis kódu
Biologické pohlaví	A	Žena
	B	Muž
Úroveň vzdělání	A	Základní
	B	Středoškolské bez maturity
	C	Středoškolské s maturitou
	D	Vysokoškolské
Rodinný stav	A	Svobodný/á
	B	V partnerském vztahu
	C	Ženatý/vdaná
	D	Rozvedený/á
	E	Vdovec/vdova
Frekvence konzumace pornografických videí bez BDSM	A	Nikdy
	B	Velmi zřídka
	C	Zhruba jednou za 2-3 měsíce
Frekvence konzumace pornografických videí s BDSM	D	Zhruba jednou za měsíc
	E	Zhruba jednou za 14 dní
	F	Zhruba jednou za týden
Frekvence masturbace	G	Několikrát za týden
	H	Skoro každý den
	I	Nejméně jednou denně

Tabulka 1 *Záznam kategorických proměnných, které charakterizují jednotlivé respondenty v našem data-setu. Kategorické proměnné označující frekvenci masturbace a frekvenci konzumace obou typů sexuálních materiálů jsou na stejné škále.*

Otázky se týkaly a) biologického pohlaví (binární proměnná), b) úroveň vzdělání (kategorická proměnná o čtyřech prvcích – tj. A-D), c) rodinný stav (kategorická proměnná o pěti prvcích – tj. A-D), d) frekvence konzumace běžné pornografie, e) frekvence konzumace pornografie s prvky BDSM, f) frekvence masturbace (vždy sedmistupňové kategorické škály – tj. A-G).

Metody

Všechny odpovědi byly kódovány jako kategorické náhodné proměnné, některé jako ordinální. Žádné z proměnných nemohou být považovány za proměnné škálové – tedy takové, které mají mezi odpověďmi shodné rozestupy (Blalock, 1960). Místo toho bylo použito kódování tzv. Kód 1 z n (anglicky one-hot-encoding – příklad níže) abychom získali vektor funkce (anglicky feature vector) ze všech odpovědí každého respondenta.

Vytvoření 38-rozměrného vektoru funkcí je popsán níže. Datový soubor k analýze tedy sestává z 38 (počet prvků k analýze) X 881 (respondentů) velké matice.

Příklad kódu 1 z n: pro otázku na "rodinný stav" existuje pět různých odpovědí (pro zamezení záměny s počitatelnými (škálovými) proměnnými převedeno na písmena A – E; Viz Tab. 2).

Segment	Počet jedinců	Věkové rozpětí	Průměr (věk)	Medián (věk)	Modus (věk)
#1	184	18–50	33.8	33.0	29
#2	45	18–47	28.4	27.0	19
#3	134	18–49	32.9	32.5	28
#4	187	18–50	32.9	33.0	32
#5	73	18–49	30.7	29.0	26
#6	52	18–50	32.6	33.5	24
#7	206	18–50	30.3	29.0	29

Tabulka 2 Rozložení věku v každém segmentu a počet členů v každém ze sedmi segmentů. Poznámka: Věk jedinců v každém segmentu je téměř shodný. Maximální fluktuace půměrů a mediánů je 10 %, proto je považujeme za podstatné pro určení členství v jednotlivých segmentech. Módy byly určeny na základě KDE rozložení věku participantů. V žádném ze segmentů není věk normálně rozložený – ve skutečnosti žádný ze segmentů ani nemá analyticky definovatelné parametrické rozložení věku.

Vektor funkce pro tuto odpověď je tedy 5-rozměrný. Odpověď "A" je nakódována jako , odpověď "B" jako , odpověď "C" jako , atd. Zjistili jsme, že tyto vektorové funkce jsou (a) ortogonální – jejich skalární součin je 0, tudíž odpověď každého respondenta je nezávislá na odpovědi jiného respondenta a (b) norma každé vektorové funkce je 1 – všechny vektorové funkce jsou normalizovány, tedy tímto překódováním do kódu 1 z n není žádná škála uměle vytvořena. Každý respondent odpověděl na všech 6 otázek.

Jako příklad si můžeme uvést respondenta, který odpověděl na první tři otázky: první odpověď B (pohlaví je binární proměnná, tedy B je druhá ze dvou možností), druhá odpověď je D (čtvrtá ze čtyř možností na otázku vzdělání – tedy nejvyšší) a třetí odpověď je C, která je třetí z pěti možností a označuje možnost vdaná/ženatý. Následný zápis do 1 z n kódu je následující: (0 1) (0 0 0 1) (0 0 1 0 0). Vektor funkce je tedy následně 11-rozměrný. . Všimněte si prosím, že vektor již není normalizovaný (norma je – protože pro příklad byly užity 3 otázky) a můžeme nalézt vnitřní souvislosti mezi jednotlivými komponenty matice. Když se podíváte na tabulku č. 1, která obsahuje dotazník, zjistíte, že celkový počet otázek je 6. Proto má normu a celkem 38 rozměrů (Tab. 3).

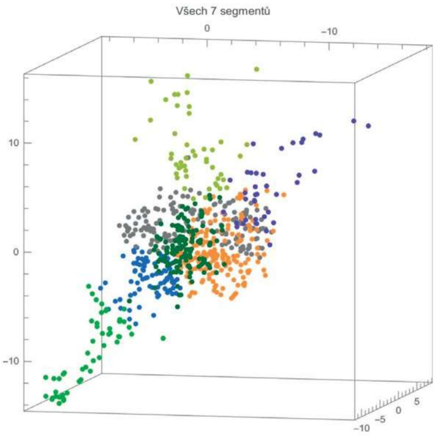
Kategorie populace	Počet obyvatel (v tisících)
A	<1
B	1–5
C	5–20
D	20–100
E	100–1000
F	>1000

Tabulka 3 Vlastnosti vztahující se k velikosti místa bydliště (vesnice, maloměsta, velkoměsta) ve kterém participanti žili. Poznámka: V kategorii „A" jsou vesnice a města s méně než jedním tisícem obyvatel a v kategorii „F" je město s více jak jedním milionem obyvatel (Praha). Detaily vztahu velikosti města a sexuálního chování jsou diskutovány v textu.

Dá se předpokládat, že obsahuje (a je jich zřejmě velmi mnoho) vnitřní souvislosti. Není třeba, aby 38-rozměrný vektor funkcí byl ortogonální s jiným 38-rozměrným vektorem funkcí.

Abychom mohli využít vzájemných vnitřních souvislostí v datech, použili jsme neuronovou síť o sedmi vrstvách - autoencodér (Mohri et al., 2018; Strang, 2019). Do analýzy vstupuje 38-rozměrný vektor funkcí a vystupuje taktéž 38-rozměrný vektor funkcí. Avšak střední vrstva má pouze 3 uzly (anglicky nods). Neuronová síť se tak snaží na každé straně symetricky tak, aby se vstup rovnal výstupu (na základě kritéria konvergence). Střední vrstva neuronové sítě o třech uzlech tak vytvoří třírozměrný vektor, který odstraní nadbytečné vzájemné vazby mezi 38 komponenty původního vektoru funkcí. Tyto třírozměrné vektory (nazývané vektory funkcí se snížením rozměrů; anglicky dimension-reduced feature vectors) jsou ve trojrozměrném prostoru (Obr. 1).

Body ve trojrozměrném prostoru, který vznikl na základě redukování rozměrů, nejsou rovnoměrně rozmístěné (Obr. 1). Následujícím krokem je nalézt shluky (klastry - segmenty) těchto bodů. Z mnoha dostupných klastrovacích algoritmů (v případě našich dat jsme vyzkoušeli 7 různých typů) jsme zvolili variantu k průměrů (anglicky K-means), kterou jsme dosáhli nejlepšího, opakovatelného, výsledku. Touto metodou bylo vytvořeno sedm klastrů (barevně odlišené v Obr. 1).



Obrázek 1 Distribuce vektorů funkcí po snížení množství rozměrů dat 881 respondentů, kteří se účastnili studie. Někteří z respondentů měli naprosto shodné rozklady funkcí, proto je v grafu pouze 667 bodů. Z vizuální analýzy distribuce dat je možné vyvodit, že nejsou uniformně rozložené a proto je vhodné tato data rozdělit do jednotlivých segmentů klastrovacím algoritmem. Příslušnost k jednotlivým segmentům je označena pomocí barevné škály, celkový počet barev (a klastrů) je 7 a toto rozdělení bylo dosaženo pomocí k-průměrů algoritmem po analýze neuronovými sítěmi. Osy a vzájemná vzdálenost bodů mezi sebou nemá žádnou výpovědní hodnotu a nelze interpretovat. Segmenty (vzniklé klastrovacím algoritmem) naopak výpovědní hodnotu mají. Skrze analýzu teplotních map (Obr. 2) lze tyto segmenty charakterizovat. V textu jsou popsány jejich konkrétní interpretace.

V tomto článku nazýváme tyto klastry segmenty.

Máme možnost reverzně ověřit, v jakém segmentu se ty které body nachází a jakým způsobem jsou rozděleny v původním 38-rozměrném vektoru funkcí a následně tak určit, k jakým konkrétním odpovědím se vztahují. Tuto metodu jsme využili k tomu, abychom mohli prezentovat (formou nestatistického sumativního vyjádření – občas hanlivě označovaného jako „počítání fazolí" [been-counting]) rozložení konkrétních odpovědí v jednotlivých segmentech (Obr. 2).

Obrázek 2 Teplotní mapy relativního rozložení odpovědí členů každého ze sedmi segmentů. Segmentace vznikla za užití klastrovacího algoritmu k-průměrů na základě rozložení rozměrově redukovaných dat od 881 respondentů (Obr. 1). Další popis je v textu.

Zajímalo nás, jak se různé atributy členů segmentů (jejich pohlaví, úroveň vzdělání atd.) spojují s populační kategorií, kde žijí (v době, kdy se dotazníku zúčastnili).

Prvním krokem bylo zkontrolovat jednotnost kontingenční tabulky (kategorie populace versus číslo segmentu) pomocí testu . Jak podrobněji rozvedeme níže, výsledek testu zdaleka není informativní. Proto používáme mnohem informativnější vyšetřovací nástroj: Korespondenční analýza (anglicky conrrespondence analysis; CA; Greenacre, 2007; Beh & Lombardo, 2014). Tento nástroj je založen na singulárním hodnotovém rozkladu (anglicky singular value decomposition; SVD; Leon, 1998; Strang, 2019) škálované kontingenční tabulky.

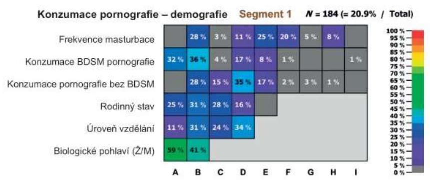
Výsledky

Pomocí algoritmu k-průměrů jsme zjistili rozložení do sedmi segmentů, v nichž se vyskytuje rozdílný počet členů – nejsou tedy velikostně srovnatelné. Tabulka 2 shrnuje počty jedinců v každém z jednotlivých segmentů. Obrázek 1 pak zobrazuje prostorové rozdělení klastrů. Pro každý klastr (segment) následně vygenerujeme teplotní mapy (anglicky heat-maps), které nám dovolují vizuálně vyhodnotit rozložení jednotlivých odpovědí v daných segmentech, jak je možné vidět v Obrázku 2.

Na základě společně se vyskytujících znaků následně charakterizujeme jednotlivé znaky společné pro celý segment:

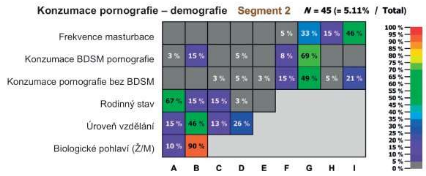
Segment č. 1 („Běžná populace") (obr. 2a): Ačkoli se jedná „jen" o zhruba 1/5 z celého vzorku, tito jedinci jsou nejlépe charakterizováni jako členové s neobvyklými sexuálními postoji v liberální společnosti. Ženy a muži jsou téměř rovnoměrně rozděleni; zhruba 1/3 má středoškolské vzdělání bez maturity a zhruba 1/3 má vysokoškolské vzdělání. Zhruba 1/3 je vdaná a málokdo je rozvedený. Konzumace pornografických videí bez prvků BDSM je z velké části mezi „téměř nikdy" a „maximálně jednou za čtrnáct dní". Konzumace BDSM pornografických videí je mezi „nikdy" (~ 1/3) a „téměř nikdy" (~ 1/3). Četnosti masturbace se vyskytují „zřídka" (~ 1/4) a jednou nebo dvakrát za měsíc (~ 1/2).

Obrázek 2a



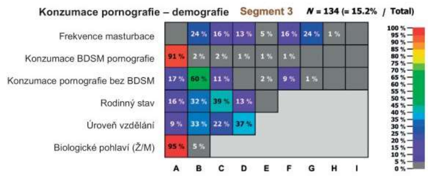
Segment #2 („Sexuálně toužící muži s nízkým vzděláním") (obr. 2b): Jedinci z tohoto segmentu tvoří ~ 1/20 vzorku, tito jedinci jsou nejlépe charakterizováni jako převážně muži s velmi vysokou sexuální spotřebou (především neinteraktivní: ~ 2/3 nejsou v partnerském vztahu). Zhruba 1/2 má středoškolské vzdělání bez maturity a zhruba 1/4 má vysokoškolské vzdělání. Spotřeba pornografických videí bez BDSM i BDSM je více než jednou týdně (the former more than daily for ~1/5 of the members). Frekvence masturbace je velmi vysoká, de facto denně až několikrát denně.

Obrázek 2b



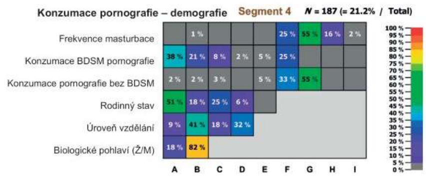
Segment č. 3 („Ženy s velmi nízkou sexuální aktivitou") (obr. 2c): Jedinci z tohoto segmentu tvoří ~ 1/7 vzorku; tyto de facto všechny ženské participantky jsou nejlépe charakterizovány jako ženy s velmi malou sexuální konzumací, které zřídkakdy sledují pornografická videa bez BDSM a nikdy se nedívají na BDSM videa. Jsou převážně v partnerském vztahu (~ 2/3); ojediněle ovdovělé (~ 1/8). Zhruba 1/3 má středoškolské vzdělání bez maturity a více než 1/3 má vysokoškolské vzdělání. Frekvence masturbace jsou rozděleny: ~ 1/2 zřídka masturbuje, zatímco ~ 1/3 masturbuje několikrát týdně.

Obrázek 2c



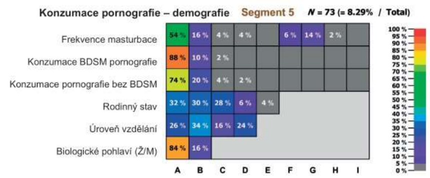
Segment č. 4 („Muži s velmi, velmi vysokou sexuální aktivitou") (obr. 2d): Jedinci z tohoto segmentu tvoří ~ 1/5 vzorku; tito de facto všichni mužští jedinci jsou nejlépe charakterizováni jako muži s velmi vysokou spotřebou videí bez BDSM, kteří nikdy nebo téměř nikdy nesledují video s BDSM. Jsou převážně bez partnerů (~ 1/2); někteří jsou ženatí (~ 1/4) Zhruba 1/2 má středoškolské vzdělání bez maturity a 1/3 má vysokoškolské vzdělání. Frekvence masturbace je velmi vysoká: více než 2/3 masturbuje jednou týdně nebo více než jednou týdně.

Obrázek 2d



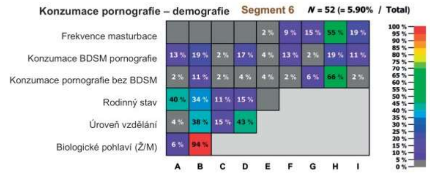
Segment č. 5 („Ženy s téměř žádnou sexuální aktivitou") (obr. 2e): Jedinci z tohoto segmentu tvoří ~ 1/8 vzorku; tyto de facto všechny ženské participantky jsou nejlépe charakterizovány jako ženy bez konzumace videí s BDSM i bez BDSM a nikdy nemasturbující (malá výjimka ~ 1/5). Jsou převážně bez partnerů (~ 1/3) nebo ve vztahu (~ 1/3) nebo jsou vdané (~ 1/4). Zhruba 1/3 má středoškolské vzdělání bez maturity a 1/4 vysokoškolské vzdělání.

Obrázek 2e



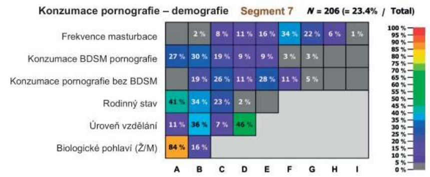
Segment č. 6 („Muži bez partnerského vztahu s extrémně vysokou sexuální aktivitou") (obr. 2f): Těchto několik málo jedinců z tohoto segmentu tvoří ~ 1/17 vzorku; tito de facto všichni mužští jednotlivci jsou nejlépe charakterizováni jako muži s velmi vysokou konzumací videí bez BDSM (2/3 denně) a velmi vysokou frekvencí masturbace (více než 1/2 denně a ~ 1/5 více než jednou denně). Jejich konzumace BDSM videí je překvapivě rozličná; téměř rovnoměrně rozložena po celé škále. Jsou převážně bez partnerů (~ 1/2) nebo ve vztahu (~ 1/3), ale jen zřídka jsou v manželském svazku. Mnoho z nich (téměř 1/2) má vysokoškolské vzdělání.

Obrázek 2f



Segment č. 7 („Ženy výjimečně s partnerským vztahem s vysokými frekvencemi masturbace") (obr. 2g): Jedinci z tohoto segmentu tvoří ~ 1/4 vzorku; tyto de facto všechny ženské participantky lze charakterizovat jako ženy s vysokým podílem vysokoškolského vzdělání (téměř 1/2), převážně bez vztahu (1/2) nebo ve vztahu, ovšem ne vdané (1/3) a s vysokou frekvencí masturbace (1/2 od čtrnáctidenní až po denní; ostatní ještě více).

Obrázek 2g

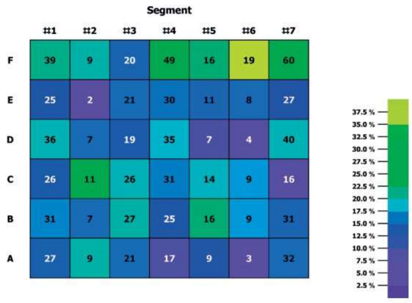


Zaznamenali jsme, že (biologické) pohlaví je ve většině segmentů nevyrovnané. Rovněž jsme také zaznamenali, že extrémní odpovědi (jmenovitě kategorie nad „G") na spotřebu pornografických videí bez BDSM a s BDSM, stejně jako frekvence masturbace, jsou zastoupeny velmi řídce, pokud vůbec.

Celkově můžeme v každém segmentu pozorovat určité snadno identifikovatelné charakteristiky jednotlivců. Tyto charakteristiky jsou často posměšně charakterizovány jako „stereotypní". Je pozoruhodné, že je bez ohledu na to, že jsou takto označeny, skutečně jde o identifikovatelné charakteristiky. Například ženy v Segmentech č. 2, 3 a 5 žijí ve vesnicích a menších městech (viz níže). Jejich postoje k sexualitě jsou poměrně omezené („konzervativní"). Mějte prosím na paměti, že ne všechny segmenty populace jsou si velikostně rovnocené – právě naopak.

Zjistili jsme, že členové v různých segmentech se vyznačují odlišnými sexuálními postoji, ať už jde o frekvenci jejich konzumace porno videí bez prvků BDSM nebo s BDSM, a/nebo frekvenci (self-reportované) masturbace. Obr. 3 ukazuje, že jejich rozdělení pozorovaných frekvencí se jeví jako nejednotné (tj. nezávislé na počtu segmentů), protože počet členů v různých segmentech se navzájem liší.

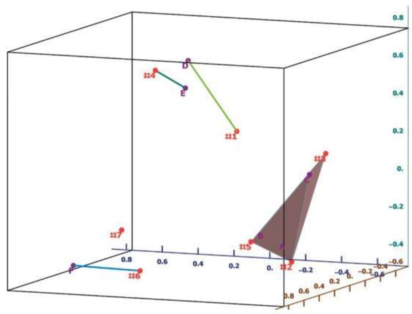
Obrázek 3



Obrázek 3 Rozložení demografických proměnných a členství v jednotlivých segmentech. Teplotní mapa ukazuje vztah mezi jednotlivými segmenty (sloupce) a demografickými proměnnými (řádky). Číslo v každé dlaždici vyjadřuje počet osob, který se v daném průniku vyskytuje. Počty zobrazené v dlaždicích jsou vyjádřením podílu segmentu, jehož členové žijí v lokalitě s danou velikostí místa bydliště. Poznámka: S ohledem na nestejnou velikost segmentů žádáme čtenáře o opatrnost v interpretaci procentuálního zastoupení.

Obr. 4 ukazuje kontingenční tabulku pro vztah mezi kategorií populace a číslem segmentu. Testem jednotnosti této kontingenční tabulky je -test (Baguley, 2012). Tento test však pouze specifikuje pravděpodobnost pozorování nehomogenního rozdělení; neodmítá nulovou hypotézu homogenity rozdělení. Ve skutečnosti je to tak, že -test nedokáže vyhodnotit, zda je zamítnutí nulové hypotézy učiněno oprávněně, tento fenomén je známý jako klam ze strany žaloby (anglicky Prosecutor's Fallacy; Spiegelhalter, 2019). „Bohužel, hodnota P je často mylně interpretována jako pravděpodobnost, že nulová hypotéza (H0) je pravdivá.

Obrázek 4



Obrázek 4 Výsledky korespondenční analýzy zjišťující vzájemné asociace mezi segmenty a velikostí místa bydliště (také viz Tab. 3) Červeně označené jsou segmenty a fialově označené jsou kategorie velikosti místa bydliště. Střepinový graf (Tab. 4 vlevo) ukazuje tři rozměry, které je třeba k vyhodnocení vzájemných asociací. Asociace byly nalezeny za pomoci algoritmu DBSCAN (Ester et al., 1996). Více informací je obsaženo v textu článku.

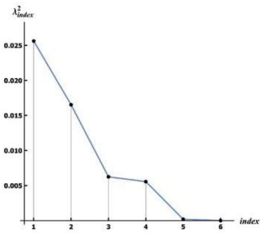
“ Tato chyba se nazývá „klam ze strany žaloby“… Malá hodnota P (například P = 0,05) nám tedy pouze říká, že v kontextu tohoto předpokladu došlo k nepravděpodobné události. “ (Krzywinski & Altman, 2017).

Kromě toho, že jsou stěží informativní, existují ještě dva další nedostatky testu : (a) výsledek testu (význam) neinformuje výzkumníka o tom, zda a jaký má nehomogenita (je-li přítomna) za následek pozorované rozdělení a (b) jaká část distribuce (rovnoměrná nebo jiná) je způsobena šumem (statistický šum, anglicky statistical noise – šum je způsoben náhodností a nevysvětlitelnými změnami v datech, jeho opakem je signál, který se snažíme v datech odhalit). Poměr šumu a signálu v datech je možné testovat. Také je možné použít metody, které nejsou šumem penalizovány, jako korespondenční analýza (CA; Greenacre, 2007; Beh & Lombardo, 2014).

Abychom demonstrovali rozdíl ve výsledku běžně užívaných metod, data jsme zanalyzovali pomocí - testu i pomocí CA, abychom zjistili, zda se členství v jednotlivých segmentech dá asociovat (nikoli korelovat) s dalšími demografickými proměnnými.

Pro -test: bylo dosaženo signifikance 0,596. Znovu upozorňujeme, že tato hodnota jednoduše uvádí, jak pravděpodobné nebo nepravděpodobné je zaznamenání tohoto rozdělení demografických kategorií, aniž by bylo možné učinit závěr o platnosti nulové hypotézy. Hodnota 0,596 tedy naznačuje, že tento výsledek není neobvyklý, pokud předpokládáme nulovou hypotézu rovnoměrného rozdělení – nic víc (Altman & Krzywinski, 2017).

Dále jsme užili singulární hodnotový rozklad na (škálované) kontingenční tabulky a získáme jejich singulární rozklad a to nakolik přispívají k druhé mocnině Frobeinovy normy (Tab. 4).



Index No.	Singular Value λ_{index}	(Singular Value)^2 λ^2_{index}	Fraction (%)	Cummulative Fraction (%)
1	0.160	0.0256	47.3	47.3
2	0.129	0.0165	30.5	77.8
3	0.0790	0.00625	11.5	89.4
4	0.0746	0.00556	10.3	99.6
5	0.0139	0.000192	0.4	100.
6	0	0	0.	100.

Tabulka 4 Výstup z analýzy SVD (Leon, 1998) ze (škálované) kontingenční tabulky (Obr. 3) Nalevo: Sutinový graf (anglicky Scree Plot) druhých mocnin hodnotových rozkladů. Vpravo: Výčet hodnotových rozkladů; jejich druhých mocnin; hodnoty jednotlivých částí Frobeiniovy normy těchto druhých mocnin; součet podílů částí Frobeiniovy normy druhých mocnin. První tři hodnotové rozklady mají nejvyšší hodnotu, téměř 90 %. Musíme vzít v potaz, že 10 % druhých mocnin Frobeinovy normy tvoří šum. Je tedy nutné vytvořit 3D graf, který nám pomůže interpretovat vztah hodnot ze (škálované) kontingenční tabulky (Obr. 4).

Uvedené druhé mocniny singulárních hodnotových rozkladů (normalizované) kontingenční tabulky nám podávají informaci o tom, že existují asociace mezi jednotlivými segmenty (sloupce) a demografickými proměnnými (řádky), jak je patrné v Obr. 4. Můžeme vyvodit, že 3 singulární hodnotové rozklady nám vysvětlují 90% signálu obsaženého v datech. Naproti tomu -test nás pouze informuje o tom, zda jsou hodnoty rovnoměrně rozloženy (respektive dostatečně rovnoměrně rozloženy). Analýza pomocí kontingenčních tabulek nás tedy jednoznačně informuje o existenci asociace mezi segmenty a jednotlivými proměnnými.

Obr. 4 demonstruje tyto souvislosti. Tyto souvislosti byly určeny za pomoci klastrovacího algoritmu DBSCAN (Ester et al., 1996) v kombinaci se (3-rozměrnými) souřadnicemi získanými ze singulárního hodnotového rozkladu. Vztahy, které jsme odhalili, jsou následující. Segment č. 1 (**„Běžná populace“**) se nejvíce pojí s velikostí města D (20 – 100 tis. obyvatel). Segment č. 4 („Muži s velmi, velmi vysokou sexuální aktivitou“) je nejvíce asociován s velikostí města E (tedy nad 100 tis. obyvatel) a segment č. 6 („Muži bez partnerského vztahu s extrémně vysokou sexuální aktivitou“) je asociovaný s velikostí města F (Praha). Segment č. 7 je („Ženy výjimečně s partnerským vztahem s vysokými frekvencemi masturbace“) se neasociuje s žádnou velikostí místa bydliště a je tedy relativně rovnoměrně rozložen mezi všechny tyto velikosti. Segmenty č. 2, 3 a 5 jsou asociovány současně s velikostí bydliště v kategorii A, B a C, tedy obecně vesnicemi a městy do velikosti 20 tis. obyvatel.

Upozorňujeme, že tyto tři segmenty (č. 2, č. 3 a č. 5) mají 252 členů (29% vzorku), ale že sexuální postoje v těchto segmentech se značně liší. Z tohoto důvodu nemusí být členové v různých segmentech považováni za jednotné ve svých sexuálních postojích, přestože žijí ve vesnicích a malých městech. Je pozoruhodné, že segment č. 2 obsahuje převážně mužské členy, kteří jsou sexuálně velmi, velmi aktivní, a to jak s ohledem na konzumaci pornografických videí, tak na frekvenci masturbace. Segmenty č. 3 a č. 5 obsahují převážně ženy, které žijí ve vesnicích a velmi malých městech. Na rozdíl od mužů, kteří žijí v těchto vesnicích a malých městech, mají pozoruhodně omezené („konzervativní“) sexuální postoje. Stereotypně formulováno, zdá se, že ženy dodržují životní styl, který, jak se zdá, prosazuje vyhýbání se sexuálním touhám, sexuální zvídavosti a aktivitním sexuálním aktivitám. Zdá se, že muži jsou sexuálně frustrovaní a v opačném chování (oproti ženám) hledají uvolnění.

Diskuze

V naší studii jsme použili novou metodu založenou na umělé inteligenci, která odhalila různé shluky charakterizující jednotlivce na základě jejich sociodemografických charakteristik a dvou druhů konzumace pornografie. Naše výsledky jsou porovnávány se zjištěními v jiných, převážně českých studiích. Je třeba zdůraznit, že česká společnost je považována za velmi otevřenou a tolerantní (v politicky liberální terminologii) v sexuálních otázkách (Widmer et al., 1998). Zjistili jsme, že v české společnosti existují značné skupiny jednotlivců, které takto charakterizovat nelze.

Této studie se zúčastnili participanti ve věku 18-50 let, kteří měli být reprezentativním vzorkem české populace, jak můžeme vidět v Tabulce 1. Žádný ze vzorků nemá participanty mladší 45 let. Obr. 4 také ukazuje, že rozdělení participantů napříč populačními kategoriemi je dosti rovnoměrné (což ukazuje i test). Jakákoli nerovnoměrnost věku (a co se týče vzdělání atd.) je proto způsobena zvláštními sexuálními atributy v každém segmentu.

Další problém: úroveň vzdělání „C" (ukončení střední školy) je v každém segmentu zastoupena nízkým procentem. Toto nízké procento je nevyhnutelné, protože ve „vyspělých společnostech" (tj. ve společnostech s dobře strukturovaným systémem terciárního vzdělávání a odpovídajícím ekonomickým základem) studenti středních škol obvykle přecházejí na vysoké školy, univerzity atd. Pouze kvůli náhodnému načasování účasti na dotazníku, kdy mohli být participanti ve fázi přechodu ze střední školy na instituce terciárního vzdělávání, dává odpověď kategorie C (ukončení střední školy) smysl. Bližší pohled na segmenty odhalí (Obr. 2), že tento fenomén nezpůsobí žádný rozdíl ve vyhodnocení dat.

Charakterizaci a popis atributů členů v různých segmentech mohou kritici nazývat „stereotypem". Zde uvedené statistické metody však dokumentují, že takové stereotypní popisy jsou nejen vhodné, ale také dobře podporovány našimi zjištěními. Skutečně zjišťujeme, že určité „typy" mužů a žen se nejpravděpodobněji vyskytují například ve vesnicích a malých městech a jiné „typy" ve velkých městech. Nabízí se otázka, zda kategorie obyvatel odvozují sociologické deskriptory, například „liberální", „konzervativní", „neprogresivní" atd. Upozorňujeme, že (a) takové sociologické kategorie zřejmě existují a projevují se ve změnách sexuálního postoje mezi segmenty; a b) sociologické postoje lze zjistit konzumací pornografických videí a rozdílnými frekvencemi masturbace v různých segmentech. Tyto klastry jsou statisticky nezávislé jednotky (a popírají tvrzení o expresi určitých rysů, které jsou normálně distribuovány v populaci; del Giudice, 2009), tedy zjištění, které by mohlo být nápomocné pro budoucí terapeutické a výzkumné práce.

Zjistili jsme, že naše statistická metodologie je velmi účinná při detekci dvou aspektů konzumace pornografie související se sexualitou a četností masturbace (self-reportované). Protože jsme analyzovali velký vzorek (881 účastníků), který má velmi dobré (téměř reprezentativní) rozložení po celé České republice, očekáváme možné podobnosti v těchto distribucích atributů, souběžně s demografickými kategoriemi, v jiných zemích se srovnatelnou infrastrukturou, dobře strukturované terciární vzdělávací systém a vyspělé (více high-tech než výrobní) ekonomiky.

Za účelem nalezení segmentů jsme využili klastrovacího algoritmu DBSCAN (Ester et al., 1996) v kombinaci se (3-rozměrnými) souřadnicemi získanými ze singulárního hodnotového rozkladu. DBSCAN je algoritmus, který pracuje bez učitele (unsupervised clustering algorithm), takže není předem definováno, kolik segmentů vznikne. Dále pak je významné zmínit, že dovoluje možnost vytváření segmentů obsahujících pouze jeden prvek – jednotlivce v našem případě. Toto je výhodou oproti alternativnímu algorytmu K-medioid, který by byl závislý na předem definovaném množství segmentů a nedovoloval by možnost vzniku výše zmíněných jednoprvkových segmentů.

Pro čtenáře může být frustrující, že při čtení tohoto textu není schopen identifikovat jednotlivé segmenty za použití teplotních map bez dopomoci umělé inteligence (Obr. 2). Jde o důsledek toho, že umělá inteligence je schopna třídit na základě mnoha kategorií současně, na rozdíl od nás – lidí – kteří jsme „od přírody" nastaveni na to, abychom se soustředili na jednu či dvě rozdělující kategorie. Algorytmy umělé inteligence jsou navíc schopny, na rozdíl od člověka a jeho přirozeného třídění, jasně a smysluplně vyfiltrovat nepodstatné a opakující se informace, které nepřináší zlepšení signálu oproti šumu.

Abychom se mohli zabývat konkrétními poznatky, které můžeme spojit s předchozími zjištěními, považujeme za důležité začít tím, že segmenty nikdy neskládají jednotlivci jednoho pohlaví. Jiné výzkumy ukazují velký přesah mezi sexuálními preferencemi, přestože výhody pro příslušníka jednoho biologického pohlaví nemusí být u druhého evidentně přítomny (Binter et al., 2013). Mezi pohlavími může existovat velmi významný rozdíl (Del Giudice, 2009), ale to nevyvozuje nulovou přítomnost druhého pohlaví v segmentu. Obě pohlaví se vyskytují ve většině našich detekovaných segmentů, konkrétně 4 ze 7 mají alespoň 15 % druhého biologického pohlaví.

Segment č. 1 obsahuje 20,9 % všech účastníků, z toho 41 % jsou muži. Participanti z tohoto segmentu mají kladný vztah k sexu a baví je pornografie, ale spíše jako inspirace (Smutny & Sulc, 2018) než jako primární prostředek uspokojení; lze je považovat za bezproblémové konzumenty (Logue, 2018). Největší segment č. 7, s 23,4 % všech účastníků, se skládá převážně z žen (84 %), které jsou spíše experimentálně založené a otevřené a sdílejí charakteristiky bezproblémové pornografické konzumace se segmentem č. 1. Jedinci v tomto segmentu však mají vysokou míru konzumace pornografie s prvky BDSM.

U dvou segmentů (č. 2 a č. 6) pozorujeme nutkavou potřebu konzumace pornografických materiálů, jak bylo také pozorováno v jiných studiích (Gaither & Sellborn, 2003; Vaillancourt-Morel, 2017; Logue, 2018). Členové segmentu č. 6 „Muži bez partnerského vztahu s extrémně vysokou sexuální aktivitou" jsou obvykle muži bez partnerského vztahu, kteří dávají přednost pornografii bez prvků BDSM. Žijí v Praze (jediné město v České republice s více než 1 milionem obyvatel) a mají vysoké vzdělání. Jejich sexuální charakteristiky tudíž nejsou podmíněny nedostatkem partnerů. Pro členy v segmentu č. 2 („Sexuálně toužící muži s nízkým vzděláním") se naopak spotřeba pornografie a vysoká frekvence masturbace jeví jako kompenzační mechanismus kvůli nedostupnosti konsensuálního partnera (Binter, 2013; Smutny & Sulc, 2018). Tito, převážně muži, mají nízké vzdělání a žijí v menších městech; takže jejich pozice na partnerském trhu není vysoká (Edlund & Sagarin, 2014). Jejich vysoká spotřeba videa s prvky BDSM naznačuje kompenzační mechanismus a online masturbace (pravděpodobná) během těchto videí slouží jako naplnění jejich sexuální touhy (Smutny & Sulc, 2018). Tyto dva segmenty, ve kterých se frekvence masturbace blíží maximu (denně nebo ještě častěji), dohromady tvoří přibližně 11 % (5,11 pro č. 2 a 5,90 pro č. 7) vzorku – ve shodě se Smutny & Sulc (2018), Mnoho dalších studií zabývajících se problematickou konzumací pornografie rovněž zjistilo úrovně přibližně 10 % (Castro-Calvo et al., 2020). Jedním z předpokládaných mechanismů může být opakované uspokojení způsobené suprnormálním stimulem. Takový „superstimulus" (Průšová et al., 2021) je založen na „vyladění" percepčního systému jednotlivce tak, aby maximalizoval zisk (Buchholz et al., 2021), když jsou mu prezentovány různé stimuly, z nichž jeden je vnímán jako „lepší než skutečný" (například nadměrnému vejci ze sádry dává pták přednost před vlastními vejci; Tinbergen, 1953); tedy chování, které je většinou maladaptivní. Zvláště frustrující by pro tyto muže mělo být sdílení sociálního prostoru se ženami ze segmentů č. 3 a č. 5., které jim nedovoluje dostatečnou (sexuální) interakci.

Segmenty č. 3 („Ženy s velmi nízkou sexuální aktivitou") a č. 5 („Ženy s téměř žádnou sexuální aktivitou") mají dva společné rysy; jedním je život v malých městech (což má za následek sníženou šanci na setkání s partnery). Další je extrémně nízká frekvence sledování pornografie. Algoritmus tak ukazuje, že Segment č. 5 se od č. 3 liší a to zejména měrou sexuální touhy, která lze odvodit z frekvence masturbace. Naproti tomu členové v segmentu č. 3 mají různou náchylnost k masturbaci (včetně jejích vysokých frekvencí). Je pozoruhodné, že segment č. 3 zahrnuje nejvyšší podíl rozvedených osob a nejnižší počet nezadaných. Domníváme se, že segment č. 3 obsahuje členy, kteří mají ze sociálních důvodů nízkou sexuální aktivitu. Možná někteří z těchto jednotlivců mohou přecházet mezi partnerskými vztahy. Obecně můžeme očekávat, že tito lidé budou mít nízkou sociosexualitu (Penke & Asendorpf 2008), pravděpodobně měli v minulosti málo sexuálních partnerů a jsou sexuálně inhibováni.

Závěr

Použití neuronové sítě překonává několik statistických problémů. Jedním z nich je „prokletí dimenzionality / prokletí vícerozměrnosti" (anglicky curse of dimensionality). Neuronová síť zmenšila rozměry z 38 na 3, aniž by eliminovala jakékoli informace (jmenovitě vzájemné závislosti mezi komponentami původního 38-rozměrného vektoru funkcí). Použití klastrových algoritmů může odhalit společné rysy mezi segmenty populace: jakými charakteristikami sexuality (frekvence masturbace, spotřeba pornografických filmů/videí a partnerský vztah – což je také indikátor sexuality!) disponují v době účasti na dotazníku.

V sekci Diskuse jsme vyhodnotili, na základě výstupů analýz, že kategorizace respondentů do jednotlivých segmentů není pouhou možností, ale že jde o rigorózně obhajitelnou metodu, která dovoluje vyvozování existence oddělených skupin (segmentů), které mají typické vlastnosti (vztahující se k jejich sexualitě), společné s ostatními členy stejného segmentu a odlišné od vlastností osob v segmentu jiném, za použití umělé inteligence.

Sexuální charakteristiky nejsou rozděleny rovnoměrně ani demograficky. Z lékařských a psychologických hledisek jsou členové v segmentech, které vykazují příliš kompulzivní sexuální chování, jako je masturbace (doprovázená konzumací pornografických videí) výrazně nad normálem. „Normální" spotřebu lze ve skutečnosti zjistit metodami AI, které zde uvádíme. Je vhodné zamýšlet se nad odbornou pomocí pro tyto jedince, s kompulzivním sexuálním chováním. Sexuologové a terapeutičtí specialisté zabývající se pomocí těmto lidem mohou získat užitečný náhled na to, jak je takové nutkavé chování ve vzájemném vztahu s frekvencemi masturbace a konzumací pornografických videí s prvky BDSM a bez prvků BDSM. Členové opačného extrému, ti s velmi potlačenou sexuální aktivitou a konzumací, mohou také potřebovat terapeutickou intervenci. Důkazy pro tuto potřebu jsou vzácné; společnost, zdá se, nepovažuje tyto lidi za ty, kteří potřebují více naplňující sexuální uspokojení.

Dále zjišťujeme, že pozorovaná stratifikace sexuality má demografický aspekt. Některé typy sexuálních postojů se vyskytují pouze ve vesnicích a malých městech, zatímco ve velkých městech jsou všechny sexuální postoje přítomné, i když v různých distribucích. Toto je, pokud je nám známo, první studie, která takové souvislosti, založené na datech, poskytuje.

Etické otázky

Projekt byl schválen etickou komisí Přírodovědecké fakulty Univerzity Karlovy (#2018/08). Všichni účastníci byli plnoletí a před účastí v online dotazníku dali souhlas k účasti ve studii. Účastníci byli upozorněni, že mohou svou účast kdykoli ukončit bez uvedení důvodu.

Střet zájmů

Autoři neprohlašují žádný střet zájmů.

Dedikace

Tento výstup je výsledkem badatelské činnosti podporované Grantovou agenturou České republiky v rámci grantu GA ČR 19-12885Y „Behaviorální a psycho-fyziologická reakce na prezentaci ambivalentních obrazových a zvukových stimulů".

Použitá literatura

Altman, N. & Krzywinski, M. (2017) Interpreting P values. Nature Methods 14, 213–214

Baguley, T. (2012) Serious Stats. A Guide to Advanced Statistics for the Behavioral Sciences. Palgrave Macmillan, Basingstoke, UK.

Beh, I.J. & Lombardo, R. (2014) Correspondence Analysis. Theory, Practice and New Strategies. Wiley, Chichester, UK.

Blalock Jr., H. M. (1960) Social Statistics. McGraw-Hill Book Company Inc., New York, N.Y., USA.

Binter, J., Leongomez, J. D., Moyano, N., Valentova, J., Jouza, L., & Klapilova, K. (2012). Sex differences in the incidence of sexual fantasies focused on evolutionary relevant objects. Anthropologie (1962-), 50(1), 83-94.

Binter, J., Boschetti, S., Hladký T., Říha, D., & Wells T. J. (in print). Vliv pornografie na přejímání emocí a sexuálního chování. Sexuologia

Binter, J., Leongómez, J. D., Moyano, N., Valentova, J., Jouza, L., & Klapilova, K. (2012). Sex differences in the incidence of sexual fantasies focused on evolutionary relevant objects. Anthropologie, 50, 83–94.

Birnbaum, G. E., & Reis, H. T. (2019). Evolved to be connected: The dynamics of attachment and sex over the course of romantic relationships. Current opinion in psychology, 25, 11-15.

Boschetti, S., Hladký, T., Říha, D. & Binter, J.,(in print). Pleasure and Cyclical States Similarities: Food Spiced by Chilli and Sex Spiced by Sado-Masochism Share Common Features, Sexuologia.

Brody, S., & Costa, R. M. (2009). Anatomy/physiology: Satisfaction (sexual, life, relationship, and mental health) is associated directly with penile–vaginal intercourse, but inversely with other sexual behavior frequencies. The journal of sexual medicine, 6(7), 1947-1954.

Buchholz, V. N., Mühle, C., Kornhuber, J., & Lenz, B. (2021) Markers of prenatal androgen exposure correlate with online sexual compulsivity and erectile function in young men. Frontiers in Psychiatry 12, 212.

Castro-Calvo, J., Gil-Llario, M. D., Giménez-García, C., Gil-Juliá, B., & Ballester-Arnal, R. (2020) Occurrence and clinical characteristics of Compulsive Sexual Behavior Disorder (CSBD): A cluster analysis in two independent community samples. Journal of Behavioral Addictions 9, 446–468.

Chivers, M. L., Seto, M. C., Lalumiere, M. L., Laan, E., & Grimbos, T. (2010). Agreement of self-reported and

genital measures of sexual arousal in men and women: A meta-analysis. Archives of sexual behavior, 39(1), 5-56.
Chivers, M. L., & Bailey, J. M. (2005). A sex difference in features that elicit genital response. Biological psychology, 70(2), 115-120.
Del Giudice, M. (2009). On the real magnitude of psychological sex differences. Evolutionary Psychology 7, 147470490900700209.
Diamond, M., Jozifkova, E., & Weiss, P. (2011). Pornography and sex crimes in the Czech Republic. Archives of sexual behavior, 40, 1037–1043.
Dines, G. (2010). Pornland: How porn has hijacked our sexuality. Beacon Press.Edlund, J. E., & Sagarin, B. J. (2014) The mate value scale. Personality and Individual Differences 64, 72–77.
Ester, M., Kriegel, H.-P., Sander, J., Xu, X. (1996). Simoudis, Evangelos; Han, Jiawei; Fayyad, Usama M. (eds.). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231.
Fedoroff, Paul J. (2008). Sadism, Sadomasochism, Sex, and Violence. Canadian Journal of Psychiatry. 53: 637–646.
Gaither, G. A., & Sellbom, M. (2003). The sexual sensation seeking scale: Reliability and validity within a heterosexual college student sample. Journal of Personality Assessment 81, 157–167.
Gray, F. (1998) At Home with the Marquis de Sade: A Life. Simon & Schuster, New York, NY, USA.
Greenacre, M. (2007) Correspondence Analysis in Practice. Chapman & Hall/CRC, Boca Raton, FL, USA.
Gyamfi, D., & Zalata, O. A. (2018). OVERDEPENDENCE OF YOUR BRAIN ON PORN. In Актуальная медицина (pp. 721-726).
Hald, G. M., & Štulhofer, A. (2016). What types of pornography do people use and do they cluster? Assessing types and categories of pornography consumption in a large-scale online sample. The Journal of Sex Research, 53(7), 849-859.
Hicks, T. V., & Leitenberg, H. (2001) Sexual fantasies about one's partner versus someone else: Gender differences in incidence and frequency. Journal of Sex Research, 38, 43–50.
Hopkins, E. (1899). The power of womanhood; or, Mothers and sons. A Book for Parents and Those in Loco Parentis. London: Wells Gardner, Darton and Co.
Krafft-Ebing, R. (2011). Psychopathia sexualis: The classic study of deviant sex. Skyhorse Publishing Inc.
Krejcova, L., Chovanec, M., Weiss, P., & Klapilova, K. (2018). 314 Gender differences in pornography consumption: a representative study in the Slovak Republic. The Journal of Sexual Medicine, 15(7), S251.
Krzywinski, M. & Altman, N. (2013) Significance, P values and t-tests. Nature Methods 10, 1041–1042
Lawner, L. (1984) I Modi. The Sixteen Pleasures. An Erotic Album of the Sixteenth Century, Peter Owen Publishers, London, UK.
Leon, S.J. (1998) Linear Algebra. 5th edition. Prentice Hall, Upper Saddle River, NJ, USA.
Logue, J. (2018) Pornography Statistics: Who uses Porn? https://www.sagu.edu/thoughthub/ pornography-statistics-who-uses-pornography. Accessed 01 June 2018.
McKeown, J.K., Parry, D.C., Light, T.P. (2018). "My iPhone changed my life": How digital technologies can enable women's consumption of online sexually explicit materials. Sexuality & Culture, 22, 340–354.
Mohri, M., Rozdamizadeh, A., Talwalkar, A. (2018) Foundations of Machine Learning (Adaptive Computation and Machine Learning Series), Cambridge, MA: The MIT Press. 504 p.
Murphy, K.P. (2012) Machine Learning. A probabilistic Perspective. The MIT Press, Cambridge, MA, USA.
Pine, J. (2013). In Bizarre Fashion: The Double-Voiced Discourse of John Willie's Fetish Fantasia. Journal of the History of Sexuality 22, 1–33.
Průšová, D., Bochetti, S., Hladký, T., Říha, D., Wells, T.J. & Binter., J. (in print) Vliv pornografie na přejímání emocí a sexuálního chování. Sexuologia
Richters, J., De Visser, R. O., Rissel, C. E., Grulich, A. E., & Smith, A. M. (2008). Demographic and psychosocial features of participants in bondage and discipline, "sadomasochism" or dominance and submission (BDSM): Data from a national survey. The Journal of Sexual Medicine 5, 1660–1668.
Sabina, C., Wolak, J., & Finkelhor, D. (2008). The nature and dynamics of Internet pornography exposure for youth. CyberPsychology & Behavior, 11(6), 691-693.
Sinha, I. (1988) Translations of The Love Teachings of Kama Sutra with Extracts from Koka Shastra Ananga Ranga and other famous Indian works on Love. Hamlyn, London, UK.
Smutny, Z., & Sulc, Z. (2018). The Internet and Consumption of Pornography. In International Conference on Internet Science (pp. 78–91). Springer, Cham.
Spiegelhalter, D. (2019) The Art of Statistics. Learning from Data. Penguin Random House, London, UK.
Steingräber, Stephan (2006). Abundance of Life: Etruscan Wall Painting. Getty Publications, Los Angeles, CA, USA. pp. 67–68, 100.
Strang, G. (2019) Linear Algebra and Learning from Data. Wellesley Cambridge Press, Wellesley, MA, USA:
Tinbergen, N. (1953). The Herring Gull's World. A Study of the Social Behaviour of Birds. Collins, London, UK.
Vaillancourt-Morel, M.P., Blais-Lecours, S., Labadie, C., Bergeron, S., Sabourin, S., & Godbout, N. (2017) Profiles of cyberpornography use and sexual well-being in adults. The Journal of Sexual Medicine 14, 78–85.
van Lankveld, J. J., Platteau, T., van Montfort, K., Nieuwenhuijs, F., & Syroit, J. (2015). The predictive validity of SIS/SES and BIS/BAS scores for sexual and non-sexual risk behavior. Personality and Individual Differences, 79, 7–12.
Weiss, P., Zvěřina, J. (2001). Sexuální chování v ČR – situace a trendy. Praha: Portál.
Weiss, P., Zvěřina, J. (2009) Sexuální chování v ČŘ: srovnání výzkumu z let 1993, 1998, 2003 a 2008 [online: http://mep.zverina.cz/files/103-sexualni-chovani-v-cr-srovnani-vy-zkumu-z-let-1993-1998-2003-a-2008.pdf]
Widmer, E. D., Treas, J., & Newcomb, R. (1998) Attitudes toward nonmarital sex in 24 countries. Journal of Sex Research 35, 349–358.
Wismeijer, A. A., & Van Assen, M. A. (2013) Psychological characteristics of BDSM practitioners. The Journal of Sexual Medicine 10, 1943–1952.

Kontakt:
prof. Dr. Hermann Prossinger
Department of Evolutionary Anthropology
Faculty of Life Sciences, University of Vienna
Vienna, Austria
e-mail: hermann.prossinger@univie.ac.at

Boschetti, S. et al. (2022). “Eye can't see the difference”: Facial Expressions of Pain, Pleasure, and Fear Are Consistently Rated Due to Chance. Human Ethology, 37, 46-72. https://doi.org/10.22330/he/37/046-072

Human Ethology 37(2022): 046-072
Research Article

“Eye can’t see the difference”: Facial Expressions of Pain, Pleasure, and Fear Are Consistently Rated Due to Chance

Silvia Boschetti[1,2], Hermann Prossinger[3], Tomáš Hladký[1], Kamila Machová[2], Jakub Binter[1,2]

[1]Faculty of Humanities, Charles University, Prague, Czech Republic
[2]Faculty of Science, Charles University, Prague, Czech Republic
[3]Department of Evolutionary Biology, University of Vienna, Vienna, Austria

jakub.binter@fhs.cuni.cz

ABSTRACT

Our research consisted of two studies focusing on the probability of humans being able to perceive the difference between faces expressing pain versus pleasure. As controls, we included: smile, neutral facial expression, and expression of fear. The first study was conducted online and used a large sample (n = 902) of respondents. The second study was conducted in a laboratory setting and involved a stress induction procedure. For both, the task was to categorize whether the facial expression was rated positive, neutral or negative. Stimuli were faces extracted from freely downloadable online videos. Each rating participant (rater) was presented with five facial expressions (stimuli) of five females and of five males. All raters were presented with the stimuli twice so as to evaluate the consistency of the ratings. Beforehand, we tested for stimuli differences using specialized software and found decisive differences. Using a Bayesian statistical approach, we could test for consistencies and due-to-chance probabilities. The results support the prediction that the results are not repeatable but are solely due to chance, decreasing the communication value of the expressions of pain and pleasure. The expression of fear was also rated due to chance, but neither neutral nor smile. Stress induction did have an impact on the perception of pleasure.

Keywords: *perception, emotion, facial expression, visual stimuli, BDSM, pain and pleasure, Dirichlet distribution, Bayesian statistical approach, Cold Pressor Task*

---

submitted: 03. Sept. 2022; revised: 01. Nov. 2022; accepted: 04. Nov. 2022

INTRODUCTION

Previous studies of stimuli assessment

The ability to estimate, or even deduce, another person's inner feelings and emotions via facial expressions (visual cues) is an essential component of human communication. This ability is particularly important when the communicated facial expression is related to danger, harm, fear, and anger on the one hand and to happiness and surprise on the other (Donato et al., 1999).

Facial expressions have high communicative value for humans; their ability to associate facial expressions with an inner state has been studied in multiple cultures (Ekman 2006; Matsumoto & Kupperbusch, 2001). These studies resulted in *Discrete Category Theory* with seven universal, expressed emotion categories: anger, contempt, disgust, fear, joy, sadness, and surprise. Each of these is considered to be the result of complex grimaces arising from distinct psychophysiological, muscular and neurological activations (Izard, 1994).

A model competing with the aforementioned model, the *Circumplex Model of Emotions*, describes emotions and the associated facial expressions as a two-dimensional phenomenon (rather than the discrete categories of the *Discrete Category Theory*); it is characterized by two perpendicular dimensions, namely valence and arousal (Russell, 1980; Posner at al., 2005). A positive value along the valence axis is often considered the result of the motivation to approach some favorable situation while, conversely, a negative valence is to avoid some unfavorable one. The intensity of an expression is along the arousal axis. The coordinates along these two dimensions are inferred from the observed facial expression; this 2D vector is the emotional assessment attribution.

These two models differ in how to characterize and distinguish each emotion; they make their predictions either by using categories or by quantifying facial expressions. These competing approaches promise straightforward differentiation possibilities. The characterization outcomes need not be equal; we pursue this issue in this manuscript.

Recent studies showed that it is very difficult to correctly identify facial expression of intense emotional states. Aviezer et al. found that the facial expression of intensive states of opposite valence, such as when a tennis player reacts to winning or losing, were not correctly identified by the raters participating in the study (Aviezer et al., 2012). Such counterintuitive observations have been supported by further recent studies (Hughes & Nicholson, 2008; Wenzler et al., 2016). It has, therefore, become evident that humans are not very good in assessing a facial expression displaying an intense emotion in the absence of some contextualization. Contextualization may include auditory (De Gelder & Vroomen, 2000), body posture (Martinez et al., 2016) or some other contextual cue (Zhou & Chen, 2009; Wieser & Brosch, 2012; Kayyal et al., 2015; Alviezer et al., 2017).

There are further influences on perceiving facial expression, such as variabilities of the raters. For example, the biological sex of the rater has been shown to systematically affect the rating of facial expressions (Rotter & Rotter, 1988; Thayer & Johnsen, 2000; Hall & Matsumoto, 2004; Hampson et al., 2006). An evolutionary argument maintains that the facial expression of anger is more easily assessed by males, as they were expected to provide preemptive protection — via anger recognition in an adversary (Rotter & Rotter, 1988). The superior ability by females in other facial expressions has been considered to be adaptive and emerged during our (human) evolutionary history. Because women are the primary care-givers, so the argument, they need to successfully assess the facial expressions of care-receivers. The biological sex of the raters of pain and pleasure expressions has been previously investigated (Hughes & Nicholson, 2008). Their results showed that pain and pleasure expressions are modulated by the biological sex and facial expression of the sender. Female raters showed a better performance in recognizing female expressions of pain. However, in the

Hughes & Nicholson study, the only facial expressions considered were pain and pleasure and no other known universal facial expression (such as fear or joy).

We note that, in previous studies, the facial expressions of differing emotions were portrayed by different male/female faces, therefore preventing investigators to explore the role of the varying expressivities of the persons expressing the emotions via their faces.

Furthermore, the inner state of the individual rating the stimulus can affect the rating. This is true for both arousal and valence (Pell, 2005; Storbeck & Clore, 2008). In several seminal studies, stress was induced by exposure to heights and situations wherein participants expected an electrical shock. A modern — more ethical, yet equally reliable — alternative is to increase physiological arousal by the Cold Pressor Task (Bullinger et al., 1984). It consists of immersing a subject's extremity into ice water for a specified period of time (Mitchell et al., 2004). The sympathetic nervous system is functionally related to the psychological concept of arousal (Dawson et al., 2000) and is responsible for mobilizing the organism's resources to meet internal physiological demands as well as those of the external environment (Salvia et al., 2012).

Studies presented in this paper

Our studies are designed to overcome some of the methodological limitations of previous studies dealing with facial expression of intense affective states.

Aim of Study I: to clarify the consistency of the assessment of such expressions, focusing on the (biological) sex of the rater as well as the biological sexes of the rated (expressers). The stimuli used were images of facial expressions with high (pain, pleasure and fear) and low (neutral and smile) intensity. Previous studies have shown that it is quite easy to recognize smile, fear and neutral and more difficult to recognize pain and pleasure. Each emotion is expressed by each male/female face (they are the two sets of 25 stimuli), so we can control for the sex and the expressivity of the rated individual.

We generate our own set of stimuli by using picture frames from videos that depict consensual acts of extreme sexual activities. We adopted a categorical methodology in which there are three ratings: positive, negative, or neutral. We predicted that, if the individual has been exposed to a negative stimulus (pain, for instance) — the grimace (with expected high intensity), will be rated as negative. In a manifestly opposite stimulus (pleasure, for instance) the rating should be positive.

Aim of Study II: to conduct a follow-up study which evaluated the impact of the inner state of the rating individual. The procedure was similar to Study I with the addition of inducing a stress (Cold Pressure Task) in order to manipulate the inner state of the raters.

METHODS

Sample

Expressers' Faces: In order to be consistent with the published terminology, we use the terms "expressers" and "faces" to describe the individuals who were shown in the 50 video frames as stimuli. We specify the biological sex of the expresser with the terms male and female. The biological sex of the expressers is evident from the video frames.

Raters: In order to be consistent with the published terminology, we use terms "expression raters" and "respondents" to describe the individuals who were presented with the stimuli and who provided their ratings. We specify the biological sex of the expresser with the terms male and female. The (biological) sex we list is the respondent's self-reported one. We deleted all ratings (n = 4) of respondents who did not report their biological sex.

In Study I: A total of 902 individuals (aged 18–50; M_{age} = 32 years, SD = 8.9 years) completed the questionnaires; 526 women (M_{age} = 30.9 years, SD = 8.3 years) and 376 men (M_{age} = 33.6 years, SD = 9.5 years). In Study II a total of 28 individuals (aged 19–30; M_{age} = 22.3 years, SD = 2.3 years) took part in the experiment; 13 women (M_{age} = 22.7 years, SD = 2.8 years) and 15 men (M_{age} = 21.9 years, SD = 1.8 years).

Criteria for inclusion were: (a) age of respondents between 18 and 50 years, and (b) at least a minimal experience with adult media, since the facial expressions used in this study were extracted from such materials.

The Two Studies

Study I: The data were collected in the Czech Republic in 2021 via the agency Czech National Panel (narodnipanel.cz) and a science-oriented online portal pokusnikralici.cz using the online platform for data collection Qualtrics®. Participants submitted responses either via computer or mobile devices (smartphones or tablets).

Study II: The data were collected in a laboratory in Prague, Czech Republic. 15 participants were presented with the same stimuli as in Study I; the lower right legs of target group members (*n* = 15) were immersed in cold water (2–4 °C) for 1½ minutes, which subsequently increased their stress level (Cold Pressor Task; CPT (Bullinger et al., 1984; Brown et al., 2017)). The control group's 13 participants' lower right legs were immersed in water at room temperature.

Stimulus Creation

A method for obtaining the stimuli, as well as their use were presented in a previously published article (Prossinger, 2021b).

From the numerous videos viewed, ten videos (five with female faces and five with male faces) were chosen. Based on the plot in each video, five frames were selected (one of neutrality, one of fear, one of pleasure, one of pain, and one with smile). Three of the authors (S.B. J.B. & T.H.) are researchers in field of human sexuality with more than 10 years of experience, specifically focusing on extreme sexual behavior and on consumption of erotic materials. All three authors (one female and two male) provided their opinion on all of the chosen videos and stimuli choices. All agreed on stimuli choice and what expression is to be expected, based on the contextual information. The agreement on stimuli choice was debated among all three researchers in dedicated meetings.

We point out that it is a common misconception that the individuals taking part in such exchanges derive sexual pleasures from pain and the two happen simultaneously. Although it is not impossible, we have found no mention of this in the published scientific literature. Rather, it should be noted that sensitivity is increased by the feeling of pain by various parts of the body (in our case mainly slapping the buttocks and thighs) and only thereafter is climax achieved. There is no doubt, due to the camera perspective, about the occurrence of the climax in male expressers. In the female expressers no such explicit method of judgement can be used, but all signals of the occurrence of climax were identified by the researchers (involving breathing, contraction of pelvic and anal sphincter muscles, facial blushing, vocalization etc.; Dubray et al., 2017), supported by self-report at the end of the video in some cases.

In each video, male/female faces expressed fear, pain and pleasure during the session, while smile and neutral facial expressions were filmed during an interview prior to the pain and pleasure experiences. All stimuli (images) presented to the raters were scaled to 600 × 600 pixels; we used triangulation between tip of the nose and pupils to ensure that the proportions of the face on the screen were comparable among the frames. No background was visible within the frames presented.

Procedures
In Study I, the set of stimuli was presented twice (Task 1 and Task 2), each time with a different randomization sequence: each stimulus appeared for 1.5 seconds at random intervals ranging from 1 to 3 seconds (so as to avoid constant/rhythmic preparedness for the stimulus presentation). Thus, for each rater, a total of 50 ratings were collected for each presentation set, a total of 100 for both presentations.

In Study II, each rater was presented with the set of stimuli (five male and five female facial expressions) only once. The reason is that the Cold Pressor Task (CPT) has limited impact on the cortisol release and this allowed us to finish the procedure within 20 minutes after the CPT ended.

Ratings
Previous literature (Robertson et al., 2010) has suggested that it is a rather challenging task to correctly identify the facial expression (e.g., to categorize the expression of fear as fear). We therefore asked our participants to rate the observed expression as either positive, neutral, or negative. We thereby avoid the problem of correct labeling and avoided any intricacies associated with a verbal categorization system. The rating was provided by using keyboard keys or a touchpad with dedicated areas with icons.

Statistical Analyses
Due to the inherent advantage of Bayesian statistics when dealing with our research questions, we implemented this approach. General descriptions follow, while more detailed descriptions, augmented by a graphical display, are provided in the Appendix.
(a)Confusion matrices: Both female and male ratings are Dirichlet-distributed (in our case: 3-parametric). The (Bayesian) method of determining whether two groups are significantly different (or not) is to calculate the confusion matrix; it is the obligatory method to use when sample sizes are small. One sample (F) has a distribution $dist_F$ and another sample (G) has a distribution $dist_G$. When there is an overlap of the *pdfs* (probability density functions) of these two distributions, a fraction of F is *TRUE* (and a fraction is *FALSE*); likewise, for G. The confusion matrix has four entries:

$$\begin{pmatrix} TRUE_F & FALSE_F \\ FALSE_G & TRUE_G \end{pmatrix}$$

If the off-diagonal elements ($\{FALSE_F, FALSE_G\}$) are small, there exists a significant difference between the distributions of F and of G (the significance level being chosen by the researcher). Observe that the sum of each row in the confusion matrix is 1 = 100%. The fractions in the confusion matrix can also be calculated using Monte Carlo methods.

(b) Possibility of effects being due to chance: In Bayesian statistics, the probability s is a random variable ($0 \leq s \leq 1$). The crucial separator for determining chance is $s = \frac{1}{2}$. The probability is either the integral of the likelihood function $\mathcal{L}(s)$ over the interval $0 \leq s \leq \frac{1}{2}$ or the integral over the interval $\frac{1}{2} \leq s \leq 1$, depending on which side of $s = \frac{1}{2}$ the mode is. In either case, the integral determines whether an observation is due to chance. (A graphical description is shown in the Appendix.) Since there are positive, neutral, and negative responses, we generate a binary case (the correct responses versus the incorrect responses); then the likelihood function is the probability density function of a Beta distribution (see Appendix). For example, for smile, the correct response is the positive rating while the neutral rating and the negative rating together are incorrect responses.

(c) A further way of testing for consistency is by estimating the likelihoods of randomly generated realizations of the two distributions and then using Wilks lambda and its χ^2-distribution in the Laplace limit of large sample sizes.

RESULTS

Correctness Probabilities and Wilks Lambda for Significant Differences

Of the five tested facial expressions, only two were correctly rated with high probability (Table 1) — the stimulus smile as a positive rating and the stimulus neutral as a neutral rating.

Table 1: The three components of the 2 × 5 = 10 modes of the 10 Dirichlet distributions of the ratings of male and female faces for the five expressions by **(a)** female raters and **(b)** male raters. (Note that each mode is the vector {$mode_A$, $mode_B$, $mode_C$}.) 'Significance' refers to the probability that the male and female face distributions are drawn from the same statistical population. Thus, a significance less than 0.05 means that the two distributions are different at the 5% significance level. **(a)** We observe that female raters rate male faces significantly differently from female faces for smile and for neutral. We also note that smile and neutral are correctly assessed, because, for smile, the mode is de facto on the *A*-axis (positive rating) and very close to 1.00, while for the neutral expression, the mode is far from both *A*- and *C*-axes (therefore $mode_B$ is close to 1.00). In all other cases, the ratings by the females are not consistent with the implied descriptions of the axes. Further discussions are in the text. **(b)** We observe that male raters rate male faces significantly differently from female faces for neutral only. We also note that neutral is correctly assessed, because, for neutral expression, the mode is far from both *A*- and *C*-axes (therefore $mode_B$ is close to 1.00). In all other cases, the ratings by the females are not consistent with the implied descriptions of the axes. Further discussions are in the text.

The probability of a correct rating (a Dirichlet distribution) ranges between 0 and 1 in all cases. The closer to 1 the result (specifically: the component of the mode) is, the higher the probability of correct identification. The remaining probabilities are distributed between the two remaining possibilities.

(a)

Female Raters (Study I)				
Stimulus	Component	Female Faces	Male Faces	Significance
	Positive	0.973	0.946	
Smile	Neutral	0.021	0.035	0.004*
	Negative	0.006	0.019	
	Positive	0.162	0.060	
Fear	Neutral	0.465	0.416	0.5
	Negative	0.373	0.524	
	Positive	0.683	0.515	
Pain	Neutral	0.000	0.000	0.3
	Negative	0.356	0.525	
	Positive	0.451	0.422	

Pleasure	Neutral	0.061	0.148	0.1
	Negative	0.488	0.430	
	Positive	0.095	0.033	
Neutral	Neutral	0.885	0.707	8×10^{-6}*
	Negative	0.020	0.286	

(b)

		Male Raters (Study I)		
Stimulus	Component	Female Faces	Male Faces	Significance
	Positive	0.938	0.930	
Smile	Neutral	0.041	0.047	0.3
	Negative	0.021	0.023	
	Positive	0.159	0.127	
Fear	Neutral	0.484	0.483	0.9
	Negative	0.357	0.390	
	Positive	0.612	0.625	
Pain	Neutral	0.000	0.010	0.4
	Negative	0.408	0.365	
	Positive	0.408	0.481	
Pleasure	Neutral	0.069	0.134	0.1
	Negative	0.523	0.385	
	Positive	0.094	0.095	
Neutral	Neutral	0.870	0.765	0.00006*
	Negative	0.036	0.140	

Smile was rated by both sexes with very high accuracy (Table 1). Male participants rated male faces correctly with 0.930 probability and female faces with 0.938 probability; female raters with 0.946 probability for male faces and 0.973 probability for female faces. In the case of female participants rating the smile stimulus, there was a significant difference in rating probability between male faces and female faces (Wilks lambda test; $P < 0.001$) with a higher probability for the latter. For the neutral stimulus, the probability of correct rating by male raters for male faces was 0.765 and 0.870 for female faces. For female participants, the probability of correct rating was 0.707 for male faces and 0.885 for female faces.

The probability of assigning a correct rating for the neutral stimulus was significantly different for male and female faces (Wilks lambda test; $P < 0.001$) in both sexes, with better ratings for female faces.

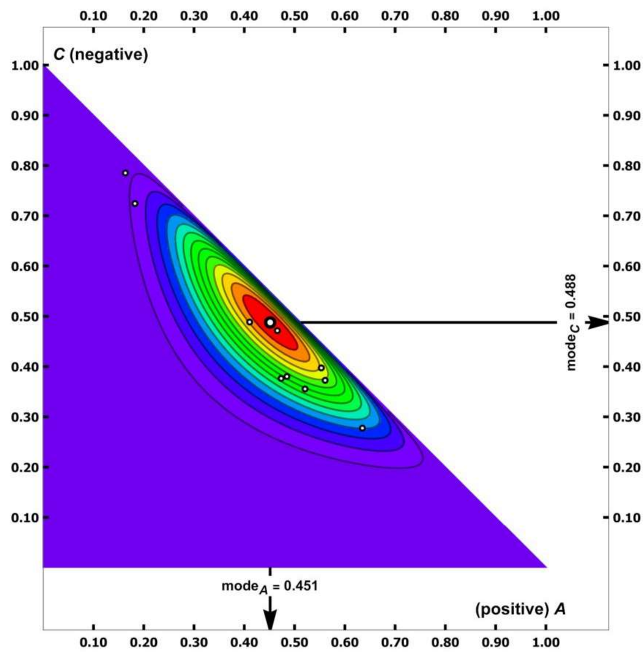
For the three other expressions (pleasure, pain, and fear), the probability of correct rating was very low (Table 1) for raters of both sexes. When rating pleasure, female raters had correctness modes of 0.422 for male faces and 0.451 for female faces; the probabilities of the rating for male faces and female faces were not significantly different; they were equally inaccurate. Male raters (when rating pleasure) had correctness modes of 0.481 for male faces and 0.408 for female faces; these were not significantly different.

When rating pain, female raters had probabilities of rating correctly of 0.525 for male faces and 0.356 for female faces; the ratings were not significantly different. For male raters, the probabilities of rating correctly were 0.365 for male faces and 0.408 for female faces; again not significantly different.

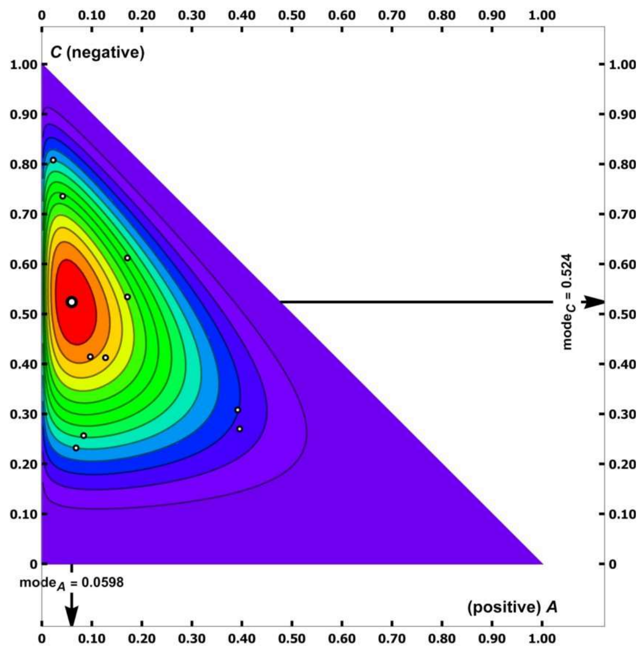
Male raters rated expressions of fear in male faces with 0.390 and in female faces 0.357 probability of correctness. Female raters' correct rating probabilities were 0.524 for male faces and 0.373 for female faces. None of these differences between the ratings of male and female faces were significantly different.

Overall, these results suggest that there is high accuracy in rating of the low arousal expressions, namely neutral and smile. There is small accuracy in the other three facial expressions. The two highly aroused facial expressions (pain and pleasure) have their ratings distributed between the two extreme ratings, namely positive and negative. This is not the case of the fear stimulus where the positive rating probability is very low and the incorrect ratings are towards the neutral rating mode.

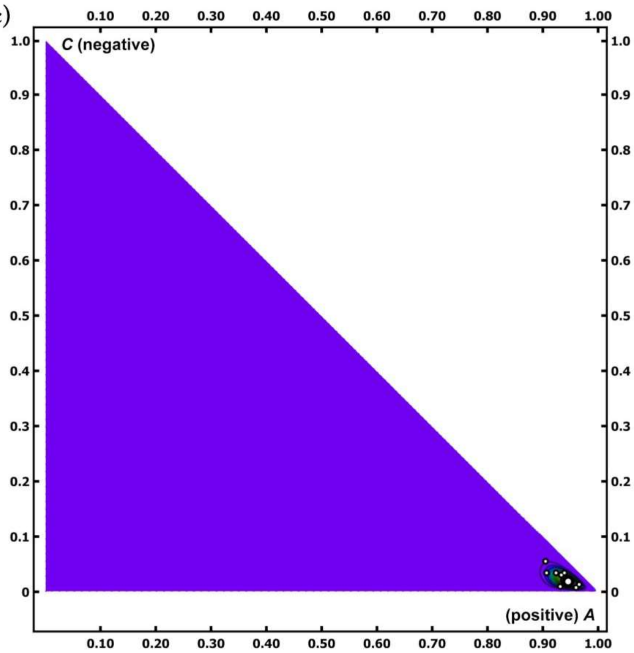
1(a)



1(b)


0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00
C (negative)
1.00 0.90 0.80 0.70 0.60 0.50 0.40 0.30 0.20 0.10 0
mode_C = 0.524
mode_A = 0.0598
(positive) A

1(c)


0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00
C (negative)
1.0 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0
(positive) A

1(d)

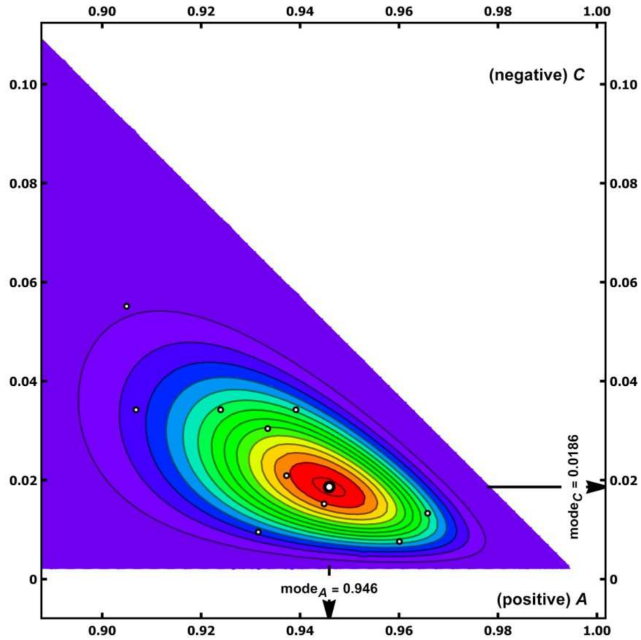


Figure 1: Graphs of three Dirichlet distributions of the ratings by 525 females: **(a)** pleasure twice expressed by females; **(b)** fear twice expressed by males; **(c)** smile twice expressed by males; and **(d)** smile twice expressed by males (detail of (c)). Here, we have chosen s_1 to be the *A*-axis (positive rating) and s_3 to be the *C*-axis (negative rating). The support of the *pdf* of the Dirichlet distribution (shown via likelihood contours) is only defined on a triangle, because $s_1 + s_2 + s_3 = 1$ (therefore, one variable — in our case *B* — cannot be rendered on an axis). The closer the mode is to the hypotenuse, the higher the value of $mode_B$. Contours are in steps of $\frac{1}{14}$ the maximum likelihood, color-coded between contours (purple lowest and red highest). The white dots with black borders are the (ten) registration sets (five faces rated twice each) and the large white dot is the mode. **(a)** We observe that the mode component for positive rating is 0.451 (very far from conventionally expected) and the mode component for negative rating is 0.488. Consequently, the mode component for neutral rating is 0.096 (close to conventionally expected). Clearly, pleasure has not been successfully rated by the 525 females; they far too often confused *A* with *C*, showing it is the result of guessing. **(b)** The mode component for positive rating is 0.0598 (as conventionally expected) and the mode component for negative rating is 0524 (very far from conventionally expected). Consequently, the mode component for neutral rating is 0.416 (extremely far from conventionally expected). Clearly, fear has not been successfully rated by the 525 females. **(c)** The mode component for positive rating is 0.973 (as conventionally expected) and the mode component for negative rating is 0.00560 (extremely low, as conventionally expected). Consequently, the mode component for neutral rating is 0.019 (very low, as conventionally expected). Clearly, smile has been successfully rated by the 525 females. **(d)** Detail of (c) showing the numerical values of the smile modes.

An ongoing discussion is about whether women are better at facial expression recognition. This is what we then tested by using a Beta distribution (a Dirichlet distribution with 2 concentration parameters). We have found only two facial expressions where the result was significant. Again, these were the two low arousal expressions (neutral and smile); but only in the case of smile did female raters show a superior ability to categorize the expression (Table 2 and Fig. 1). We note that this difference is negligible since ratings by both sexes were highly accurate (0.935 vs 0.965).

Table 2: The modes of the (Beta) distributions of male and female ratings of the stimuli, separated by sex of the faces. The columns labeled pdtc show the probabilities due to chance. Low probabilities (in the table: $P < 0.001$ — a significance level very far below the conventional $P = 0.05$) show that only the modes of the ratings of stimuli neutral and smile are not due to chance, and these are highly significant. All other modes are due to chance and therefore uninterpretable. The column $sig_{(male\ vs\ female)}$ shows whether the probability that the two (Beta) distributions of male and female facial expressions are significantly different.

Stimulus	Raters	$Mode_{female}$	pdtc	$Mode_{male}$	pdtc	$sig_{(male\ vs\ female)}$
Smile	male	0.924	<0.001	0.930	<0.001	ns
	female	0.935	<0.001	0.965	<0.001	< 0.002
Fear	male	0.369	>0.999	0.358	>0.999	ns
	female	0.458	>0.999	0.756	>0.999	ns
Pain	male	0.361	>0.999	0.411	>0.999	ns
	female	0.395	>0.999	0.383	>0.999	ns
Pleasure	male	0.463	>0.999	0.409	>0.999	ns
	female	0.412	>0.999	0.445	>0.999	ns
Neutral	male	0.683	<0.001	0.839	<0.001	< 0.001
	female	0.665	<0.001	0.834	<0.001	< 0.002

Heat Maps of individual expresser rating frequencies

Fig. 2 and Table 3 show fractions of participants of both sexes who correctly rated the facial expressions by the individual expressers. The interesting outcome is that there are expressers who are overall better (i.e. inducing a correct rating with a higher probability) than other expressers; furthermore, being better (in the above sense) is not uniformly distributed over all facial expressions.

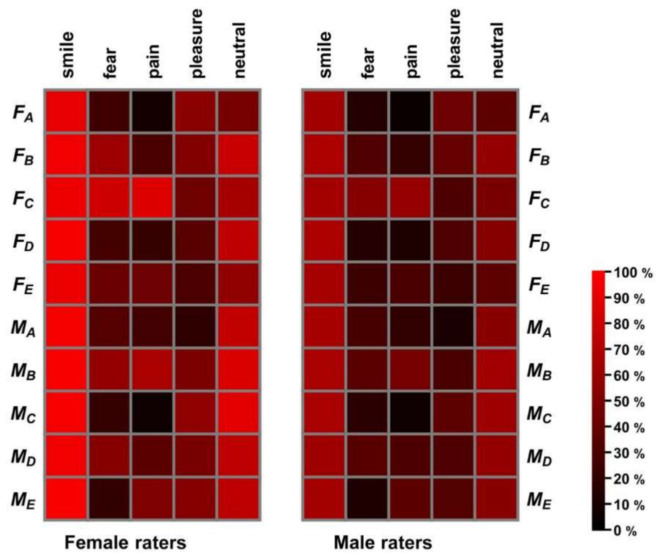


Figure 2: Two heat maps showing the correctness probabilities of ratings by males and females of the male faces and the female faces, face by face. The male faces and female faces are labeled $\mathbf{F_{index}}$ and $\mathbf{M_{index}}$. We observe that female raters correctly rate the expression smile in all female faces and all male faces with very high probability; males somewhat less than females. Remarkably, $\mathbf{F_C}$ and $\mathbf{M_B}$ were rated by the females with a higher correctness probability for fear and pain than were all other female faces and male faces. This phenomenon is comparable, with a lower correctness probability, for the male raters. The females rated all male faces and all female faces with a high probability (70–95%) of correctness for the expression neutral.

Consistencies between Task 1 versus Task 2

Since we presented all stimuli as two consecutively presented tasks, each in a different randomized order, we have the possibility to test the consistency of the ratings. To do so, we have used a Bayesian probability test; the ratings (correct versus incorrect) are Beta distributed. We found one significant result (Table 4), namely only when female raters rated male faces.

Table 3: The modes of the ratings of the Beta Distributions of male and female raters of the five female faces (prefix 'f') and the five male faces (prefix 'm') expressing the labeled facial expressions during Task 1 and Task 2. The modes are for the probability *s* being correct; if the postulated rating is to be 'negative', then *s* is the probability of the raters rating the facial expression as negative.

	Male raters		**Female raters**		**Male raters**		**Female raters**	
	Task 1	**Task 2**	**Task 1**	**Task 2**	**Task 1**	**Task 2**	**Task 1**	**Task 2**
Expresser	*Smile*				*Fear*			
fA	0.912	0.910	0.905	0.939	0.725	0.594	0.754	0.772
fB	0.955	0.910	0.945	0.932	0.655	0.511	0.782	0.757
fC	0.918	0.912	0.933	0.907	0.928	0.914	0.973	0.946
fD	0.957	0.939	0.966	0.960	0.291	0.322	0.406	0.440
fE	0.920	0.904	0.924	0.937	0.733	0.734	0.810	0.764
mA	0.930	0.931	0.968	0.945	0.670	0.639	0.597	0.610
mB	0.941	0.936	0.971	0.983	0.926	0.898	0.938	0.921
mC	0.944	0.960	0.971	0.977	0.500	0.390	0.456	0.397
mD	0.880	0.904	0.945	0.949	0.788	0.774	0.791	0.786
mE	0.915	0.960	0.977	0.968	0.245	0.327	0.269	0.281
Expresser	*Pain*				*Pleasure*			
fA	0.069	0.098	0.090	0.109	0.705	0.630	0.659	0.589
fB	0.458	0.412	0.459	0.548	0.709	0.758	0.680	0.622
fC	0.924	0.917	0.954	0.940	0.522	0.427	0.520	0.372
fD	0.238	0.278	0.330	0.340	0.502	0.492	0.395	0.411
fE	0.455	0.478	0.485	0.527	0.373	0.356	0.355	0.349
mA	0.297	0.187	0.284	0.211	0.166	0.163	0.201	0.172
mB	0.734	0.756	0.739	0.755	0.471	0.728	0.560	0.696
mC	0.087	0.117	0.059	0.079	0.564	0.551	0.601	0.582
mD	0.524	0.583	0.440	0.509	0.526	0.432	0.557	0.457
mE	0.571	0.664	0.543	0.605	0.523	0.432	0.594	0.497
Expresser	*Neutral*							
fA	0.524	0.612	0.470	0.534				
fB	0.816	0.824	0.787	0.814				
fC	0.681	0.758	0.654	0.755				
fD	0.752	0.773	0.743	0.764				
fE	0.527	0.566	0.570	0.557				
mA	0.765	0.777	0.762	0.812				
mB	0.901	0.899	0.859	0.922				
mC	0.883	0.883	0.886	0.930				
mD	0.824	0.840	0.743	0.840				
mE	0.769	0.846	0.736	0.848				

Ratings due to chance

One of the benefits of the Bayesian analytical approach is the possibility to test whether the result obtained is consistent ('real' in common parlance) or if it is obtained due to chance. The probability of the result being due to chance (columns pdtc in Table 2) ranges between 0 and 1; the closer to 1, the more probable that the result is due to chance. The closer the mode (columns $Mode_{sex}$ in Table 2) is to $\frac{1}{2}$, the higher the probability (not: likelihood) of the result being due to chance. For all our results, we obtained extreme ends of the possible outcomes only. Specifically: the two low-arousal expressions (neutral and smile) are not due to chance with probability pdtc < 0.001 (Table 2). In other words, the results are not due to chance at all. A completely opposite result was obtained for the case of the high-arousal faces (fear, pain, and pleasure). The probability is > 0.999, therefore almost certainly due to chance.

Table 4: The consistency of ratings between Task 1 and Task 2. A high significance means that the ratings in Task 1 and Task 2 are consistent. Only one rating (by female raters of male facial expressions — highlighted in light gray) was significantly different between Task 1 and Task 2.

Stimulus	Raters	$Face_{female}$	$Face_{male}$
Smile	male	> 99%	> 99%
	female	44%	> 99%
Fear	male	> 99%	> 99%
	female	> 99%	> 99%
Pain	male	> 99%	> 99%
	female	> 99%	> 99%
Pleasure	male	> 99%	> 99%
	female	> 99%	17%
Neutral	male	> 99%	> 99%
	female	> 99%	< 1%

Note: The significance of the result is reported in the last two columns.

These results provide further evidence for the above mentioned results, specifically the results of rating the high-arousal stimuli. Not only are the ratings spread between (typically two) options (positive and negative in case of pain and pleasure and negative and neutral in case of fear) but these ratings are also due to chance.

Stress-induced rating differences

In Study II, we analyzed the differences in the distributions of ratings in the two groups of participants (control versus stressed). The confusion matrices (Table 5) display the probabilities of differences of the ratings by the two groups of participants (separately for male faces and female faces). At a 10% significance level (Caelen, 2017), only two off-diagonal elements are significantly different: male faces expressing smile and expressing pleasure were more accurately rated by the stressed group. The shifts (not shown) in correct rating are not due to chance at 5% significance level. (For the *CDF* of the Beta Distributions, the conventional 5% significance level is applicable.)

Table 5: The confusion matrices (entries in %) between the distributions of the ratings by the controlled and the stressed raters, separated by male faces versus female faces. At a significance level of 10% (Caelen, 2017), only two distributions are significantly different (male faces expressing smile and male faces expressing pleasure). A confusion matrix calculation was used instead of the statistical machinery of Wilks lambda, because the sample sizes ($n_{control}$ and $n_{stressed}$) were far from the Laplace condition.

Stimulus	Male Faces	Female Faces
Smile	$\begin{pmatrix} 91.6 & 8.4 \\ 9.7 & 90.3 \end{pmatrix}$	$\begin{pmatrix} 76.3 & 23.7 \\ 18.2 & 81.8 \end{pmatrix}$
Fear	$\begin{pmatrix} 64.3 & 35.7 \\ 30.8 & 69.2 \end{pmatrix}$	$\begin{pmatrix} 75.1 & 24.9 \\ 22.4 & 77.6 \end{pmatrix}$
Pain	$\begin{pmatrix} 57.2 & 42.8 \\ 33.6 & 66.4 \end{pmatrix}$	$\begin{pmatrix} 51.5 & 48.5 \\ 35.1 & 64.9 \end{pmatrix}$
Pleasure	$\begin{pmatrix} 91.5 & 8.5 \\ 7.4 & 92.6 \end{pmatrix}$	$\begin{pmatrix} 84.8 & 15.2 \\ 13.1 & 86.9 \end{pmatrix}$
Neutral	$\begin{pmatrix} 71.2 & 28.8 \\ 27.1 & 72.9 \end{pmatrix}$	$\begin{pmatrix} 73.7 & 26.3 \\ 23.9 & 76.1 \end{pmatrix}$

DISCUSSION

Stimuli Selection

There are multiple ways to produce stimuli for testing. Most often, trained actors or actresses are asked to produce facial expressions that are later rated by professionals or naïve respondents in a pre-test. Whenever the within-rater agreement is sufficiently high, the stimulus was used for testing (an example of this procedure has been published in Kätsyri & Sams, 2008). In our case, this approach is not possible for two reasons: (a) because our hypothesis postulates that the two expressions that are of highest interest to us (pain and pleasure) are putatively indistinguishable, asking pre-test raters to distinguish these would not be sensible, and (b) using stimuli labeled during pre-test as pleasure or pain would inherently lead to testing whether participants agree on representations of pain and pleasure (that is to say, whether there is a common mental representation as discussed by Chen et al., 2018). The ethological validity of such a result would be extremely limited and has been criticized in a recent publication (Van Der Zant & Nelson, 2021). Instead, we followed the methodology of one of the pioneering articles on the topic (Aviezer et al., 2012). The authors searched the internet and chose the stimuli (facial expressions of tennis players) based on the context of winning or losing. The context was not known to their raters but the authors were certain about the outcome of the match and therefore which valence (positive or negative) the stimulus represents.

Following this methodology, we searched for videos online, only including webpages that allowed a free download option. As an extension to the previously mentioned article (Aviezer et al., 2012), we went one step further in our stimuli choice and insisted on finding (and using) individuals of both biological sexes expressing all the five desired expressions.

Ethics of using isolated video frames as stimuli

Although the affliction of pain in the context of extreme, yet consensual, sexual activities is typically due to the actions of some other person, the individual experiencing it knows that he/she consented beforehand and has the ability to demand termination at any time while the scenario evolves; he/she thus retains a high degree of control. Typically, a 'safe-word' is used to terminate such physical and/or psychological activities, and, subsequently, after termination, nurturing behaviors are then provided to such an individual. Pain itself is not an aim of the behavior but rather the goal of increasing sensitivity and priming for greater pleasure (in a sexual sense). It is rare that injuries (apart from bruises) are inflicted on the pain-receiving individual. As stated on the production webpage, all participants in the video clips were informed (not by us, but by the directors) about the to-be-filmed scene contents; they agreed to participation and were interviewed by members of the production team after the scene was completed.

We consider the use of such stimuli as beneficial for science (granted: these stimuli are perceived as controversial by some). They offer the possibility of novel understandings about the problems of the perceptions of facial expressions in several of the evolutionarily most relevant contexts. The acquired knowledge (some of which we have obtained and are presenting in this paper) can, and will certainly be, used in the fields of education, sexuality-related prevention, law enforcement, and therapy. We therefore maintain that the benefits by far outweigh the objections to using such stimuli.

Novelties

Our studies confirmed the results of previous research about the facial expressions of affective states with high arousal — in the absence of further contextual clues. Specifically, the human ability to distinguish between positive and negative valence in cases of facial expressions of extremely high arousal is very weak (Aviezer et al., 2012). Our six design innovations (novelties) allowed us to provide further insights into this topic.

First, because every rater rated each stimulus twice, we could test for consistency. We discovered an increase in accuracy with the second presentation of low arousal stimuli; the increase is ascribable to a recall effect, a learning effect, and a familiarity effect. In the cases of high arousal stimuli, on the other hand, the two ratings appear to be consistent (no difference in accuracy); there cannot be any significant differences, however, because these ratings are due to guessing.

A second novelty is to find the probability of a result being due to chance. To do so, we use Bayesian statistics, which is particularly useful for this challenge. For low arousal stimuli, we find that the result is not due to chance; this infers there must be some mechanism and repeatability is to be expected. In contrast, the three high arousal stimuli (fear, pain and pleasure) are due to chance with an extremely high probability. We find that this is valid for both sexes; we must therefore conclude that discussing any sex-differences is meaningless.

The third novelty is avoiding potential (statistical) noise effects. Every expresser displayed all five facial expressions: fear, smile, pain, pleasure, and neutral.

The fourth novelty is one of design: it deals with every expresser presenting the same five facial stimuli to all male and female raters. We thereby improved (we claim) the reliability of statistical interpretability. Of the biases in data collection (confirmation bias and selection bias), we avoided the latter in this way.

The fifth novelty is the testing of expressions of pain and pleasure not only by women, but also by men.

The sixth novelty deals with a possibility of comparing the ratings of the facial expression of pleasure by males with those by females. The relation between the putative inner feelings of a male when he expresses pleasure facially is much less questionable than for a female. This novelty results in an important departure from the methodology adopted in many other studies of ratings of facial stimuli.

Responses

There have been numerous publications dealing with: (a) differences in facial expression production and degree of expressiveness in various cultures, as well as (b) comparisons of expressiveness of neurotypical versus neurodivergent individuals. Recent publications (Barrett et al., 2019) have concentrated on the individual differences in facial expression production by neurotypical individuals within one culture. We use heat maps (Fig. 2) to display, for the first time, the evidence that raters rate the individual expressers with varying probabilities of success. In other words, we register that some expressers are rated with a higher accuracy for more than one expression whereas others for only one expression.

This phenomenon of some expresser being more accurately rated than others warrants future research, as other circumstances may influence the expressivity of an individual. In naturally occurring (uncontrolled) situations, the strength of the stimulus needed to trigger an affective response varies among individuals. The subsequent research issue is to what degree this affective response triggers a corresponding facial expression. Therefore, the use of naturalistic stimuli results in highly uneven expressions, which necessitate the application of Bayesian statistics.

We did not expect the ratings of fear to be distributed almost equally between negative and neutral. Previous publications provide an explanation: of all the so-called basic expressions, the fearful expression is the least recognizable one, because it is brief and oftentimes admixed with other ones.

Pain and pleasure ratings are almost equally distributed between the extremes positive and negative, with very few neutral ratings. Our results for pain and pleasure ratings are different from those in previous publications.

Human Ratings versus AI Ratings

Our results show that there are no differences between the ratings of pain and pleasure — when rated by humans. We note, however, that there are objective methods for detecting a quantifiable difference in muscle configurations associated with different facial expressions. The recent increase in computation power coupled with the progress in artificial intelligence (AI) techniques provides appropriate tools to test the pain/pleasure rating differences objectively. So there is a difference, but it remains undetectable by humans. In contrast to the publications relying on FACS (Aviezer et al., 2012), we show, in a recently published study (Prossinger, 2021a), how to use an alternative method, based on AI image analysis, to detect objective differences. This algorithmic approach was used to distinguish fearful from neutral faces with a high success rate (Prossinger et al., 2021a). These findings support the existence of an actual difference between two facial expressions. The differences in the expressions are indeed present (and detectable with AI methods) but human raters were unable to detect them with sufficient accuracy.

An interesting comparison, using the same stimuli as in this paper, is provided in another study (Prossinger et al. 2021b); it evaluated the precision of distinguishing stimuli categories. The algorithms found significantly different categories in the case of female expressers. An extension of this research was recently published (Prossinger et al., 2022) with a larger number of female expressers experiencing pain and pleasure. This publication derives important implications about how clustering relates to human raters' inabilities to reliably distinguish between the expressions of pain and pleasure. The study precisely enumerated the far-from-trivial steps necessary for correct classification, which cannot be expected from human vision uncalibrated towards a single individual. There are four clusters and two isolates. These clusters were detected after noise removal. The discovery of the necessity of noise removal provides further support for the two main arguments about the human inability to correctly rate the differences between pain and pleasure. First, the inter-individual facial expression variations are considerable. Second, the (intra-individual) noise component in each specific perception is high. Consequently, it is possible that humans can fine-tune their perception towards certain individuals, especially socially close ones (partners, other family members, or colleagues, for example) thus putatively mitigating noise interference. The AI algorithms are, as shown above, able to overcome this problem.

It is important to point out the inter-individual variability. Even though healthy individuals are all equipped with facial muscles essential for basic emotion expression and the variability of the muscles involved is minimal (Waller et al., 2008), there are many influences related to the uniqueness of each individual´s expressions and limitations in their identification for other individuals in real world scenarios. Some of these limitations are: the fact that people choose or need to wear spectacles, some have beards, some are adorned with jewelry or expensive makeup. All may obstruct or alter the assessment of the facial expression.

Further complications may arise in individuals who experienced facial nerves-related disorders or other central nervous system damage. Furthermore, expressers' age-related features, their fat distribution, their skin texture, their general degree of facial expressiveness, and the morphology of their facial muscles are known to impact the production of their facial expressions (and consequently the probability of correct identification). Therefore, angles and distances between the facial features and their changes from neutral to expressive states constitute the individual expresser's identity (Kande et al., 2000; Yi et al., 2014). We took advantage of this identity uniqueness by using AI algorithms, because the facial identity is rather consistent in adult individuals and it allows for such human-computer interaction (Cohn et al., 2007). Possibly, individual expression familiarity potentially increased the accuracy of correct expression estimation by other humans if they are exposed to an individual for an adequately long, yet unknown in extent, period of time.

In the two studies involving the human raters that are presented in this paper, familiarity was expressly excluded. An interesting next step would be to test such a proposed explanation. Previous studies within this familiarity framework have been conducted on sadness, anger, and happiness; the results are mixed (Zhang & Parmley, 2015). In children, research on pain vocalizations has been published (Corvin et al., 2022); it claimed that learning is the mechanism for obtaining proficiency with respect to specific expressers. It would be worthwhile to compare how successful individuals are in assessing (rating) their partners and relatives in extremely (non-sexually) arousing moments (such as in sports encounters) with the ratings of strangers' facial expressions.

Influence of Arousal on Ratings

A further factor that influences the ratings is that the perception can be affected by the inner state of the rater. We rarely stay calm when encountering highly arousing situations such as winning or losing, reunion with family members, or sexual interaction — in striking contrast to common rating assessment tasks in a research context. These situations involve emotional coupling and affect mirroring and cause a dynamic attribution process (Hasson & Frith, 2016). Indeed, the state of the rater affects the perception of the expresser.

Dutton and Aron (1974) made their participants rate ambiguous pictures with the Thematic Apperception Test (TAT). Those with induced anxiety rated the situations in the pictures as having increased sexual connotations. If we assume that pain-pleasure is equally ambiguous as the stimuli (pictures) used in the TAT, we would predict a shift towards a more positive rating for both pain and pleasure. Brown et al., (2017) were among the first to design a comparable test by using the Cold Pressor Task to find a possible shift towards the negative rating in the case of a surprised face — which was considered ambiguous by those authors. Three facial expressions had been presented: happy face, surprise face, and angry face. The happy face and the angry face were not affected by the induced stress. Those stressed rated the surprise face as more negative.

Our Study II is a more refined version of Brown et al.'s study, because we used five facial expressions comprising 50 stimuli. Also, our analysis is based on Bayesian statistics, which avoids sample size issues and allows for further insights, notably due-to-chance probabilities.

The outcomes of our Study II, in which we manipulated the arousal of the raters, document no shift in rating at 5% significance level. The observed shift towards more positive rating only happened in the cases smile and pleasure in male expressers at 10% significance. The choice of significance levels in confusion matrices is dynamic; research indicates that the choice 5% is rarely warranted; 10% is to be preferred (Caelen, 2017).

Implications

Facial expressions of pain, pleasure and fear are uninformative. Because the display (of pain and pleasure) is ambiguous, the signal perceived by the rater is uninformative. The misinformation can be exacerbated by arousal change in the rater.

We therefore consider an implication to be: verbal communication is a practical resolution of the above ambiguities during many (but perhaps not all) interactions in real life.

LIMITATIONS AND FUTURE DIRECTIONS

One seeming limitation is the prediction of null results. In a statistical sense, it is considered problematic to test for a null effect (but that is perhaps due to Null Hypotheses Statistical Testing conventions and the associated fallacies). Bayesian statistics is not susceptible to such a problem (because the method does not violate Bayes' Theorem) and specifically includes testing for a null result. Therefore, this approach is promising for future research.

We tested for two types of null result. One null result (often observed): the outcome of a statistical test shows that the observed effect is due to chance. The other type we tested for: that the observed difference of a result that is not due to chance but the detected difference is valid with a very small probability.

In both studies presented here, the samples of both stimuli and raters consisted of members of a Caucasian population, since the diversity of population in the Czech Republic is minimal. The results, although very strong, may not be directly generalizable to other populations.

Female sexual pleasure is difficult to assess; but this difficulty applies to all related research. There are claims that even self-reports would not be sufficient. Devices used for measuring female sexual arousal are insufficiently reliable (Meston et al., 2004; Cooper et al., 2014; Meston et al., 2019), so we cannot rely on their applicability in this investigation. As in other studies that attempt to relate arousal with female pleasure expression, we use the pragmatic approach: for stimulus creation, it is sufficient to adopt the convention of relying on using already existing, freely downloadable videos. Researchers who question this pragmatic approach must then reject the validity of a vast number of studies dealing with facial expression of pleasure, not only those using videos. However, it should be pointed out that applying the AI methods to facial expressions (Prossinger et al., 2022) have the potential of resolving this impasse.

By the same token, we feel the need to address the possibility that the expression seen does not match the inner feeling of the expresser. This is not a design flaw but involves an inherently biological aspect in the field of research using naturalistic stimuli.

Furthermore, the expression of fear as a reliable stimulus may be considered problematic since the expressers were aware of the fact that, ultimately, the situation is safe: no permanent damage is de facto guaranteed. Fear, of all expressions considered basic, has the lowest identification reliability rate, and this is especially true in naturalistic expression scenarios. In other words, the results obtained are less unusual than may appear at first glance.

Lastly, it should be pointed out that the situation of sexual play is not transferable to other types of interaction where such mismatches can be found, e.g., sport, fighting, injury infliction. Therefore, generalization of our findings to such fields should be used with caution.

CONCLUSIONS

Recent studies dealing with facial expressions are shifting from laboratory-produced situations with pre-tested expressions towards the more real-world relevant way of stimuli creation in order to obtain context-dependent facial expressions. Due to the change in study design, the observed outcomes are remarkably different. These different outcomes challenge many of the cornerstones of this research field. Furthermore, our study is currently unique in repeatedly using one expresser's face for all five expressions. In other words, participants were presented with 50 individual stimuli, one at a time, of five different individuals expressing five different grimaces (namely smile, fear, pain, pleasure, and neutral). We find that, for human raters, perception of facial expressions of pain and pleasure are ambiguous. The participants were specifically requested to supply a categorical rating so as to avoid errors related to descriptive ratings. All insights were obtained by using a Bayesian statistical approach, which also allows for testing probabilities due to chance and a reliability measure for a null result.

The results for low-arousal expressions (smile and neutral) confirm that the method and the analytical approach are appropriate for investigating the observations. The low-arousal expressions were rated with high accuracy and with high probability, inferring that these are repeatable results. Our findings regarding high-arousal expressions, on the other hand, confirm that, even though there are objective differences in the expressions of pain and pleasure (which were tested using AI methods), they are indistinguishable by humans, especially when trying to ascertain such differences in strangers.

The rating options for pain and pleasure are actually binary (positive or negative). Even when offering a neutral distractor, we find that the ratings are always due to chance. Furthermore, this result is repeatable, which we tested by two presentations of each stimulus. In other words, guessing is the only reason for a null rating — and we did not find a learning effect. This disqualifies further analyses regarding statistical variabilities among the raters. Similar results were obtained for the fear expression. Thus, the ratings were predominantly distributed between the negative and neutral category; rarely was fear rated as positive. However, all these ratings were also due to chance.

To our knowledge, this is the first time that the ambiguous facial expressions of one expresser were presented to participants who were also in a condition of increased arousal. This procedure shifted ratings to more accurate ones, we found, namely for two positive facial expressions (smile — a low arousal expression — and pleasure — high arousal expression) of male expressers. The other expression ratings were unaffected by arousals induced in the raters. This could suggest that there is, with arousal, a selective shift in the positive expression perception: it is in concordance with the original work on misattribution of arousal.

ETHICS

Even though the materials presented to the participants were not *per se* of a sexual nature (as only facial expressions were presented) we made precautions to limit any negative impact on our participants.

Informed consent
In Study I: An online information text and consent form was supplied; after reading it, a box was to be ticked by each participant (indicating their informed consent) prior to their participation.

In Study II: Two informed consent forms were to be manually/personally signed. The first was presented to a potential rater prior to participation; it included all the information about procedures (including the CPT), safety measures, kinds of data collected, and risks. The second informed consent form consisted of a full disclosure of the aim(s) of the study, the expected impact of the procedures, and the possible implications for the rater signing this second form. It was to be signed after the debriefing procedure (see below). If the second consent form was not signed, the collected data was discarded (and therefore not used in the analysis).

Post-study Support and Debriefing
All parts of the design and debriefing were conducted in co-operation with a trained psychologist who also supervised all data collection.

For Study I we supplied the participants with a list of contacts: (1) to the principal investigator, (2) to a psychological counseling center, and (3) to an organization that deals with sexuality-related issues.

During the debriefing phase for Study II, every rater participated in a debriefing discussion by a trained psychologist directly after the completion of data collection. The rater then received a written detailed description, with a full explanation of the possible negative aspects of the experiment, especially those related to the stress-induction procedure, and was also supplied with a list of contacts: (1) to the principal investigator, (2) to a psychological counseling center, and (3) to an organization that deals with sexuality-related issues.

ACKNOWLEDGMENTS
The study was preregistered on OSF portal: https://osf.io/bhk6m/.

FUNDING
This research was funded by the Czech Science Foundation in the project with number GACR 19-12885Y and titled “Behavioral and Psycho-Physiological Response on Ambivalent Visual and Auditory Stimuli Presentation”.

INSTITUTIONAL REVIEW BOARD STATEMENT
The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Faculty of Science, Charles University, Prague, Czech Republic (Protocol Code 2018/08, approval date: 2 April 2018).

DATA AVAILABILITY STATEMENT
The data are available on the OSF portal.
The frames were extracted from commercially available online videos. As the videos are proprietary, we can only make the extracted frames we used available from the corresponding author (upon reasonable requests originating from a serious institutional email address).

CONFLICTS OF INTEREST
The authors declare no conflict of interest.

REFERENCES

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science, 338*(6111), 1225–1229. DOI

Aviezer, H., Ensenberg, N., & Hassin, R. R. (2017). The inherently contextualized nature of facial emotion perception. *Current Opinion in Psychology, 17*, 47–54. DOI

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest, 20*(1), 1–68. DOI

Brown, C. C., Raio, C. M., & Neta, M. (2017). Cortisol responses enhance negative valence perception for ambiguous facial expressions. *Scientific Reports, 7*(1), 1–8. DOI

Bullinger, M., Naber, D., Pickar, D., Cohen, R. M., Kalin, N. H., Pert, A., & Bunney Jr, W. E. (1984). Endocrine effects of the cold pressor test: relationships to subjective pain appraisal and coping. *Psychiatry Research, 12*(3), 227–233. DOI

Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence, 81*(3), 429–450. DOI

Chen, C., Crivelli, C., Garrod, O. G., Schyns, P. G., Fernández-Dols, J. M., & Jack, R. E. (2018). Distinct facial expressions represent pain and pleasure across cultures. *Proceedings of the National Academy of Sciences, 115*(43), E10013–E10021. DOI

Cohn, J. F. (2007). Foundations of human computing: Facial expression and emotion. In: *Artifical Intelligence for Human Computing* (pp. 1-16). Springer, Berlin, Heidelberg. DOI

Cooper, E. B., Fenigstein, A., & Fauber, R. L. (2014). The faking orgasm scale for women: Psychometric properties. *Archives of Sexual Behavior, 43*(3), 423-435. DOI

Corvin, S., Fauchon, C., Peyron, R., Reby, D., & Mathevon, N. (2022). Adults learn to identify pain in babies' cries. *Current Biology, 32*(15), R824–R825. DOI

Dawson, M., Schell, A., Filion, D. (2000). The electrodermal system. *Handbook of Psychophysiology*. Cambridge University Press. DOI

De Gelder, B. & Vroomen, J. (2000) The perception of emotions by ear and by eye. *Cognition & Emotion 14*, 289–311. DOI

Donato, G., Bartlett, M.S., Hager, J.C., Ekman, P., Sejnowski, T.J. (1999). Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*(10), 974–989. DOI

Dubray, S., Gérard, M., Beaulieu-Prévost, D., & Courtois, F. (2017). Validation of a self-report questionnaire assessing the bodily and physiological sensations of orgasm. *The Journal of Sexual Medicine, 14*(2), 255-263. DOI

Dutton, D. G., & Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology, 30*(4), 510. DOI

Ekman, P. (2006). *Darwin and facial expression: A century of research in review.* NY: Academic Press

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behavior, 27*(6), 401–416. DOI

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion, 4*(2), 201. DOI

Hasson, U., & Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1693), 20150366. DOI

Hughes, S. M. & Nicholson, S. E. (2008). Sex differences in the assessment of pain versus sexual pleasure facial expressions. *Journal of Social, Evolutionary, and Cultural Psychology* *2*(4), 289. DOI

Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research. *Psychological Bulletin, 115*(2), 288–299. DOI

Kanade, T., Cohn, JF, & Tian, Y. (2000). Comprehensive database for facial expression analysis. *Proceedings fourth IEEE international conference on automatic face and gesture recognition* (cat. No. PR00580), pp. 46-53. DOI

Kätsyri, J., & Sams, M. (2008). The effect of dynamics on identifying basic emotions from synthetic and natural faces. *International Journal of Human-Computer Studies, 66*(4), 233–242. DOI

Kayyal, M., Widen, S., & Russell, J. A. (2015). Context is more powerful than we think: Contextual cues override facial cues even for valence. *Emotion, 15*(3), 287. DOI

Matsumoto, D., & Kupperbusch, C. (2001). Idiocentric and allocentric differences in emotional expression, experience, and the coherence between expression and experience. *Asian Journal of Social Psychology, 4*(2), 113–131. DOI

Martinez, L., Falvello, V. B., Aviezer, H., & Todorov, A. (2016). Contributions of facial expressions and body language to the rapid perception of dynamic emotions. *Cognition and Emotion, 30*(5), 939–952. DOI

Meston, C. M., Levin, R. J., Sipski, M. L., Hull, E. M., & Heiman, J. R. (2004). Women's orgasm. *Annual Review of Sex Research, 15*(1), 173–257.

Meston, C. M., & Stanton, A. M. (2019). Understanding sexual arousal and subjective-genital arousal desynchrony in women. *Nature Reviews Urology, 16*(2), 107–120. DOI

Mitchell, L. A., MacDonald, R. A., & Brodie, E. E. (2004). Temperature and the cold pressor test. *Journal of Pain, 5*(4), 233–237. DOI

Pell, M. D. (2005). Nonverbal emotion priming: evidence from the 'facial affect decision task'. *Journal of Nonverbal Behavior, 29*(1), 45–73. DOI

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17*(3), 715–734. DOI

Prossinger, H., Binter, J., Hladký, T., & Říha, D. (2021a). Using Neural-Network-Driven Image Recognition Software to Detect Emotional Reactions in the Face of a Player While Playing a Horror Video Game. *International Conference on Human-Computer Interaction* (pp. 258–265). Springer, Cham. DOI

Prossinger, H., Hladky, T., Binter, J., Boschetti, S., & Riha, D. (2021b). Visual Analysis of Emotions Using AI Image-Processing Software: Possible Male/Female Differences between the Emotion Pairs "Neutral"–"Fear" and "Pleasure"–"Pain". *The 14th PErvasive Technologies Related to Assistive Environments Conference* (pp. 342–346). DOI

Prossinger, H., Hladký, T., Boschetti, S., Říha, D., & Binter, J. (2022). Determination of "Neutral"–"Pain","Neutral"–"Pleasure", and "Pleasure"–"Pain" Affective State Distances by Using AI Image Analysis of Facial Expressions. *Technologies, 10*(4), 75. DOI

Roberson, D., Damjanovic, L., & Kikutani, M. (2010). Show and tell: The role of language in categorizing facial expression of emotion. *Emotion Review, 2*(3), 255–260. DOI

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior, 12*(2), 139–148. DOI

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161. DOI

Salvia, E., Guillot, A., & Collet, C. (2012). Autonomic nervous system correlates to readiness state and negative outcome during visual discrimination tasks. *International Journal of Psychophysiology*, *84*(2), 211–218. DOI
Storbeck, J., & Clore, G. L. (2008). Affective arousal as information: How affective arousal influences judgments, learning, and memory. *Social and Personality Psychology Compass, 2*(5), 1824–1843. DOI
Thayer, J., & Johnsen, B. H. (2000). Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. *Scandinavian Journal of Psychology, 41*(3), 243–246. DOI
Van Der Zant, T., & Nelson, N. L. (2021). Motion increases recognition of naturalistic postures but not facial expressions. *Journal of Nonverbal Behavior, 45*(4), 587–600. DOI
Waller, B. M., Cray Jr, J. J., & Burrows, A. M. (2008). Selection for universal facial emotion. *Emotion, 8*(3), 435. DOI
Wenzler, S., Levine, S., van Dick, R., Oertel-Knöchel, V., & Aviezer, H. (2016). Beyond pleasure and pain: Facial expression ambiguity in adults and children during intense situations. *Emotion, 16*(6), 807. DOI
Wieser, M. J. & Brosch, T. (2012) Faces in context: a review and systematization of contextual influences on affective face processing. *Frontiers in psychology* 3, 471. DOI
Yi, J., Mao, X., Chen, L., Xue, Y., & Compare, A. (2014). Facial expression recognition considering individual differences in facial structure and texture. *IET Computer Vision, 8*(5), 429-440. DOI
Zhang, F., & Parmley, M. (2015). Emotion attention and recognition of facial expressions among close friends and casual acquaintances. *Journal of Social and Personal Relationships, 32*(5), 633–649. DOI
Zhou, W. & Chen, D. (2009). Fear-related chemosignals modulate recognition of fear in ambiguous facial expressions. *Psychological Science, 20,* 177–183. DOI

APPENDIX

Dirichlet Distribution

The ratings by female raters are Dirichlet distributions (in our case with three concentration parameters $\{a_A, a_B, a_C\}$), as are those of the males. We predicted the repeats (Trial I versus Trial II) to be the same, and we tested for that. We therefore have, for female raters rating five female faces displaying fear, ten registration sets with triples $\{n_A, n_B, n_C\}$ in each set, with $n_A + n_B + n_C = 10$. The *pdf* (probability density function) of the Dirichlet distribution $\mathcal{Dir}$, called the likelihood function $\mathcal{L}(s_1, s_2, s_3) = pdf(\mathcal{Dir}(a_A, a_B, a_C), s_1, s_2, s_3)$ with concentration parameters $\{a_A, a_B, a_C\}$ and probabilities s_1, s_2, s_3 of observing the variables var_1, var_2, var_3 is

$$\mathcal{L}(s_1, s_2, s_3) = pdf(\mathcal{Dir}(a_A, a_B, a_C), s_1, s_2, s_3) = \frac{\Gamma(\alpha_A + \alpha_B + \alpha_C)}{\Gamma(\alpha_A)\Gamma(\alpha_B)\Gamma(\alpha_C)} s_1^{\alpha_A - 1} s_2^{\alpha_B - 1} s_3^{\alpha_C - 1}$$

with $s_3 = 1 - s_1 - s_2$ and $0 \leq s_i \leq 1\ \forall i\ i = 1 \ldots 3$; $\Gamma(\cdots)$ is the Gamma function.

The two modes for *A* and *C* are $mode_A = \dfrac{(\alpha_A - 1)}{(\alpha_A + \alpha_B + \alpha_C - 3)}$ and $mode_C = \dfrac{(\alpha_C - 1)}{(\alpha_A + \alpha_B + \alpha_C - 3)}$.

If we are interested in axes *A* and *B*, rather than *A* and *C*, then the formulae are cycled. Below, we explain why we use which axes and when. Note that the formulae for the modes are straightforward, suggesting we need not use the (somewhat complicated) formula for the probability density function *pdf*. However, there are no closed algebraic formulas for the uncertainty intervals for the *pdf* of the Dirichlet, but there are contours of uncertainty (see, for example, Fig. 1), and these contours are oddly-shaped smooth curves. We need to analyze the contour geometry in order to interpret possible overlap (which enables us to determine whether the rating distributions of male and female raters are significantly different).

Bayesian estimation of guessing

Each face is rated as exhibiting one of the five facial expressions. We do not expect, but do postulate — as a test — that the facial expression smile (for example) will be rated positive, while the facial expression pain will be rated negative. We use a Bayesian approach to determine the maximum likelihood of a correct probability(!). For each face of each facial expression rated by the females (say), let n_1 be the number of ratings that agree with the postulated rating, while n_2 is the number of ratings that disagree with the postulated rating (then $n_1 + n_2 = n$; $n = 526$ for female raters; $n = 376$ for male raters). In Bayesian statistics, in which the probability *s* is a random variable, the likelihood function, for this situation, $pdf(s) = \mathcal{L}(s)$ of *s* is a Beta Distribution

$$\mathcal{L}(Be(\alpha, \beta), s) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha - 1}(1 - s)^{\beta - 1} = \frac{\Gamma(n_1 + n_2 + 2)}{\Gamma(n_1 + 1)\Gamma(n_2 + 1)} s^{n_1}(1 - s)^{n_2}$$

The probability (in Bayesian statistics) of observing a result disagreeing with the postulate is then,

$$\int_0^{1/2} \mathcal{L}(Be(\alpha, \beta), s)ds$$

The most likely probability s_{ML} is the mode. s_{ML} = mode = $\frac{\alpha - 1}{(\alpha - 1) + (\beta - 1)}$. We note that the postulate is always *s*, even if the postulated rating is negative (as in the case of pain).

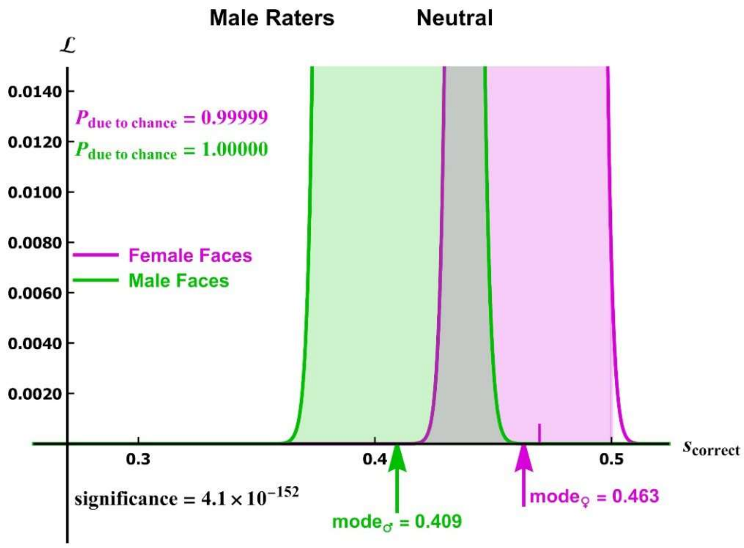


Figure A-1: An example of the relationships between the independence of two samples and the determination of an outcome due to chance. The likelihood functions of the Beta distributions of the two samples are shown, along with the areas under the curve (shaded); these areas show the probability of the observed distribution of the ratings being due to chance (i.e. guessing by the male raters of the presented stimulus). One is $P_{\text{due to chance}}$ = 1.000 … (very high probability that the raters are guessing), the other is $P_{\text{due to chance}}$ = 0.99999 … (again, very high probability that the raters are guessing; in this latter case, the small unshaded area clarifies why the probability is close, but not equal to, 1.0000…). The modes are significantly different, because the probability that the two observed rating sets are drawn from the same statistical population is 4.1×10^{-132}; in other words, it is extremely unlikely that the two distributions are samplings from the same statistical population. This significance has been calculated with Wilks lambda (see below). The peaks of the likelihood functions are not shown; the details near the s_{correct}-axis have been shown.

An example of a result is shown in Fig. A-1. For each rating set (male or female) of all 10 faces, we obtain, for each expression, two modes, one for Task 1 and one for Task 2.

Testing for independence of two distributions: Wilks Lambda

Given: two samples of ratings (of the stimulus pleasure, say), one by females (total counts n_F with n_G correct) and one by males (total counts n_M with n_H correct). The distributions are Beta distributions. If $s = s_{\text{correct}}$ is the probability of a correct rating, then the likelihood function is, for the females,

$$\mathcal{L}_F(s) = \frac{\Gamma(n_F + 2)}{\Gamma(n_H + 1)\Gamma(n_F - n_G + 1)} s^{n_G}(1 - s)^{n_F - n_G}$$

and, for the males,

$$\mathcal{L}_M(s) = \frac{\Gamma(n_M + 2)}{\Gamma(n_H + 1)\Gamma(n_M - n_H + 1)} s^{n_H}(1 - s)^{n_M - n_H}$$

We generate random numbers N_A using the female likelihood function and random numbers N_B using the male likelihood function. We then estimate the ML Beta distribution $dist_A$ using the N_A random numbers and the ML Beta distribution $dist_B$ using the N_B random numbers. We also estimate the ML distribution $dist_{AB}$ of the combined random numbers $N_{AB} = N_A \cup N_B$. We calculate the three log-likelihoods

$$\text{logLike}_A = \ln\mathcal{L}(dist_A|N_A)$$
$$\text{logLike}_B = \ln\mathcal{L}(dist_B|N_B)$$
$$\text{logLike}_{AB} = \ln\mathcal{L}(dist_{AB}|N_{AB})$$

and Wilks Lambda

$$\Lambda_{Wilks} = -2(\text{logLike}_{AB} - (\text{logLike}_A + \text{logLike}_B)),$$

which is, in the Laplace/frequentist paradigm (i.e. *sample size* → ∞), χ^2-distributed with $df = df_{AB} - (df_A + df_B)$ degrees of freedom (df = 2 in our analysis). Thus, if significance = CDF($\chi^2(2)$, Λ_{Wilks}) is very small, we conclude that the probability that the two samples with their respective Beta Distributions are improbably drawn from the same statistical sample.

Fig. A-1 shows an example. The significance is very small, so we conclude that the probability that the two samples (male and female faces rated by male raters) are drawn from the same statistical population is significantly small; therefore, the modes are significantly different. However, both the distributions of rating samples cannot be excluded from being due to chance.

Apoidolia: A New Psychological Phenomenon Detected by Pattern Creation with Image Processing Software together with Dirichlet Distributions and Confusion Matrices

Silvia Boschetti[1[0000-0002-8048-4062]], Jakub Binter[1,2[0000-0001-5304-2130]], and Hermann Prossinger[3[0000-0002-1114-7687]]

[1] Faculty of Science, Charles University, Prague, Czech Republic
[2] Faculty of Social and Economic Studies, University of Jan Evangelista Purkyně, Ústí nad Labem, Czech Republic
[3] Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria
silvia.boschetti@natur.cuni.cz

Abstract. Traditionally, a strict dichotomy between belief and science is commonplace; recently, scientists investigating this complex phenomenon have designed studies to more thoroughly approach this dichotomy. Cognitive and perceptual processes developed under evolutionary pressures, especially stressful and harsh conditions ones, result in an enhanced perception of patterns. This has led to the emergence of a bias to perceive visual patterns that are manifestly random, a phenomenon called pareidolia ('overperception'), whereas the bias of assigning meaningfulness to general random patterns is apophenia. Methodological constraints in visual perception studies have led researchers to focus only on false perception — pareidolia. We extended the methodology by using image processing software to generate random maps and random maps of varying transparency overlying nonrandom patterns to identify not only false perception but also lack of perception. We used several questionnaires. For every query in every questionnaire, we classified two groups: those participants who never made a mistake versus those who made a mistake (which would be either perceiving a pattern where there is none or not perceiving an existing pattern) at least once. We then estimated the two groups' Dirichlet distributions of the responses, and calculated the confusion matrix to find significant differences. The Rational Experiential Inventory yielded a significant differentiation between the two groups. In addition to perceivers with pareidolia, we found that some perceivers failed to identify an existing pattern. We call this psychological phenomenon apoidolia ('underperception') — seeing no pattern when there is one. To our knowledge this is the first time this psychological phenomenon has been empirically detected.

Keywords: Apoidolia, Apophenia, Pareidolia, Dirichlet Distribution, Confusion Matrix.

15 Introduction

In recent years, the number of scientific studies of religion and supernatural beliefs has increased considerably, primarily due to novel approaches, novel methodologies and novel experimental procedures. One approach to study such phenomena is the cognitive study of religion, which focuses on those cognitive functions and mechanisms that support religious thinking and beliefs [1]. These mechanisms may have an evolutionary adaptive function in specific contexts and may therefore shape perception and interpretation of the perceived world [2]. In order to ensure survival, humans must rely on the ability to detect patterns and correctly identify their meaning; only then are they able to correctly/optimally respond [3]. This process of pattern identification is, however, subject to errors and mistakes. The error management theory (EMT) postulated that a false positive error (detecting a pattern where there is none) has an evolutionary advantage over a false negative error (not detecting a pattern where there is one), especially during stressful conditions. Indeed, the risk of not perceiving a dangerous animal (typically an almost camouflaged snake) is much higher than the cost of perceiving such an animal when it is actually not there [4]. The perception of patterns and 'meaningful' interconnections among elements of these — even when they are actually not present (i.e., illusory pattern perception) is called apophenia [5], and is named pareidolia when such phenomena arise in the context of visual stimuli [6]. In studies conducted on a non-clinical population, pareidolia was associated with different types of supernatural beliefs (including religious beliefs, beliefs in coincidence, as well as beliefs in conspiracy, and also magical beliefs) [7, 8]. Indeed, individual differences in perceptual processes during the elaboration and interpretation of external stimuli could work as substrate for religious and supernatural beliefs [9].

Previous studies mainly focused on false recognition of faces [7, 10] — face pareidolia —, which is seeing the presence of structure in objects with a collection of patterns that resemble the elements of (for example) a face; consequently, participants would therefore identify such patterns as images of faces. Since a face constitutes a very specific type of stimulus, with high evolutionary relevance for humans, a specific brain area is dedicated to processing these stimuli [6, 11]. The results obtained using such stimuli may not be repeatable when studying general illusory pattern perceptions (apophenia), however.

Our study aims to investigate the errors in pattern perception and identification without restricting the results to specific adapted stimuli (i.e. faces) and to clarify the relationship of such perceptual errors with beliefs and thinking styles.

We constructed our own stimuli by repeatedly using a multidimensional random number generator available in MATHEMATICA (from Wolfram Technologies®), to produce random maps (in which, therefore, no discernible patterns were present and the entropy was maximized) and then underlying them with identifiable patterns — concretely with geometric shapes such as octagons and triangles. Such shapes are common in nature [12]. We used these geometric figures as patterns in order to avoid biases involved in the perception of biological objects (such as leaves). We controlled the transparency of the random maps to manipulate the difficulty of perception and thereby the ambiguity of the stimulus.

Using these types of stimuli, we could investigate not only the presence of false positive errors in perception (pareidolic perception) but also for the presence of false negative errors (which we call apoidolic perception). To our knowledge, we are the first to identify and statistically evaluate both, pareidolic and apoidolic perception in a non-clinical population.

16 Materials and Methods

16.1 Participants

Our sample consisted of 105 participants from the Czech Republic, aged between 18 and 50 years ($mean \pm SD = 33.7 \pm 7.7$ years); 67 were women ($mean \pm SD = 33.6 \pm 7.5$ years) and 38 men ($mean \pm SD = 33.8 \pm 8.1$ years). The participants were recruited from a science-oriented web community and the data were collected online.

16.2 Stimuli

We used a multidimensional random number generator to produce random maps of colored squares (henceforth called random patterns) and underlying them with several geometrical objects (Fig. 1). This method ensures a decrease in entropy as transparency is increased. We present three types of stimuli for each series: one in which the geometrical figures were fully covered with the random pattern (No Pattern condition), one

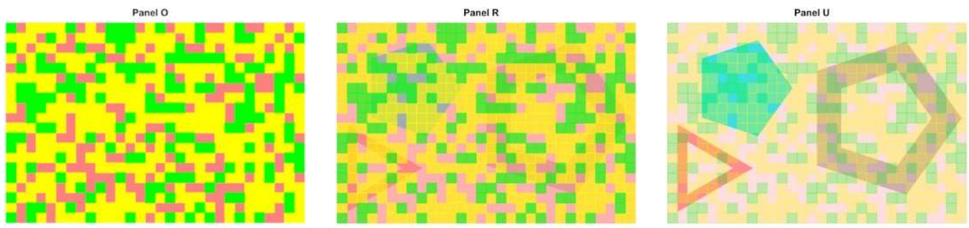


Fig. 1. Three of the patterns displayed to the participants. Panel O: the random pattern (No Pattern condition); Panel R: sample of the random pattern with underlying figures partially revealed (Partial condition), and Panel U: sample of the random pattern with underlying figures almost completely revealed (Reveal condition). The x- and y-coordinates of the colors (from a sample of a color triple) were randomly generated; the colors of the squares were also randomly chosen. The transparency of the random pattern was increased in seven steps from opacity (Panel O) to almost complete transparency (Panel U). The colors of the geometric objects were not drawn from the color triples, nor were they changed when increasing the transparency of the random overlay pattern.

in which the transparency of the random map was increased, allowing for the geometrical figures to be partially visible (Partial condition) and a third image in which the transparency had been further decreased and the geometrical figures were well identifiable (Reveal condition) while the (almost transparent) random map remained partially visible. We constructed three series of stimuli; each series contained one stimulus with three conditions (No Pattern, Partial and Reveal) for a total of nine experimental stimuli presented to every participant. To avoid learning bias during the data collection, we mixed the experimental stimuli with filler stimuli.

16.3 Questionnaires

We used a suite of questionnaires (listed below) to determine different types of supernatural and religious beliefs and thinking styles. The questionnaires were translated into Czech by translators with psychological training; their reliability was confirmed by back-translation.

The Questionnaire on Coincidence [13] measures perceived coincidence experiences. This questionnaire has two parts; Part A measures the occurrence of episodes classified as different types of coincidence, while Part B investigates possible explanations for the occurrence of coincidences. Due to the type of study we were conducting and the analysis we chose in this paper, we currently used only Part A.

The Religion Commitment Inventory (RCI-10) [14] tests religious beliefs — especially the commitment to a religious group or community.

The Illusory Beliefs Inventory (IBI) [15] measures magical thinking and has three subscales to identify separate components: (a) magical beliefs (general beliefs in magic), (b) spirituality (beliefs in a higher power) and (c) thought-action-fusion (beliefs that thoughts can shape reality).

The Rational/Experiential Multimodal Inventory (REIm) [16] measures the rational and experiential thinking style as two separate ways to process information. The questionnaire consists of four subscales, one scale for rationality and three separate scales for different aspects of the experiential thinking style: intuition, emotionality and imagination.

16.4 Analytical Approach

We inventoried pattern recognition responses as either correct or erroneous. Depending on the stimulus presented, the error could be a false positive error (a pattern was identified when no pattern was present, as when the No Pattern condition had been presented) or a false negative error (no pattern was identified when a pattern was present, as when the Partial condition and the Reveal condition had been presented). In order to avoid the 'Lady Tasting Tea' fallacy/error [17], several random patterns and several degrees of transparency were used.

For numerous reasons, we decided not to compute a questionnaire index but to use a query-by-query approach to explore the relevance of the actual items (query scores) of the questionnaires.

A Bayesian approach was used. For each query in each questionnaire, two heat maps were constructed: the one with no errors and one with at least one error (Fig. A-1). For each query, the response frequencies are Dirichlet-distributed with five concentration parameters (the questionnaires used 5-option Likert scales). As described in the Appendix: for each query, overlap of the *pdf*s (probability density functions) of the two Dirichlet distributions, one for correct pattern detection versus one for incorrect pattern detection (either pareidolia or apoidolia) occurs. The overlap can be used to construct the confusion matrix; the significance level we adopt is 10% (the level for confusion matrices equivalent to 5% in conventional significance tests [18]).

17 Results

It should be highlighted that — as opposed to traditional classical test theory — we used the query-by-query approach, because all the steps are based on an underlying mathematical theory. We describe the difference (at 10% significance level [18]) between subjects that correctly identified the stimuli and subjects that committed an error separately for each condition: No Pattern condition (no pattern was present) versus Partial condition (the geometrical figures were partially visible) versus Reveal condition (the geometrical figures were well identifiable).

Table 7. Summary of significant queries for each questionnaire. The table lists the number of significant queries per questionnaire and their fraction of the entire questionnaire (in %). * indicates 80% significant queries (or higher).

Questionnaires	**False positive** (No Pattern)	**False negative** (Partial)	**False negative** (Reveal)

Coincidence (7 queries)	5 (71%)	6 (86%)*	4 (57%)
RCI (10 queries)	2 (20%)	2 (20%)	5 (50%)
IBI-Magical Beliefs (10 queries)	3 (30%)	2 (20%)	10 (100%)*
IBI-Spirituality (8 queries)	4 (50%)	8 (100%)*	8 (100%)*
IBI- TAF (5 queries)	3 (60%)	2 (40%)	3 (60%)
REIm-Rationality (12 queries)	12 (100%)*	11 (92%)*	12 (100%)*
REIm-Emotion (10 queries)	8 (80%)*	9 (90%)*	10 (100%)*
REIm-Imagination (10 queries)	8 (80%)*	9 (90%)*	9 (90%)*
REIm-Intuition (10 queries)	9 (90%)*	9 (90%)*	10 (100%)*

For the Coincidence questionnaire (Part A): in the No Pattern condition, five of the seven queries were significant (queries number 1, 2, 3, 6, and 7); in the Partial condition, six queries were significant (queries number 1, 2, 4, 5, 6, and 7), while for the Reveal condition four queries were significant (queries number 1, 2, 4, and 6).
For RCI-10 (10 queries): in the No Pattern condition, two queries were significant (queries number 1 and 6); in the Partial condition, only two queries were significant (queries number 5 and 10), while for the Reveal condition five queries were significant (queries number 2, 4, 5, 6, and 9).
For the subscale Thought-Action-Fusion (5 queries): in the No Pattern condition three queries were significant (queries number 1, 3 and 4); in the Partial condition two queries were significant (queries number 1 and 5), and three in the Reveal condition (queries number 2, 3 and 4).
For the subscale Magical beliefs (10 queries) of the IBI: in the No Pattern condition three queries were significant (queries number 1, 6 and 8); for the Partial condition two queries were significant (queries number 5 and 6), while for the Reveal condition all 10 queries were significant. For the subscale Spirituality (8 queries): in the No Pattern condition four queries were significant (queries number 2, 3, 4, and 8); in both the Partial and Reveal condition all eight queries were significant.
For the subscale Rationality of the REIm: in the condition No Pattern all 12 queries were significant; the same results were obtained for the Reveal condition; for the Partial condition 11 queries were significant (the only exception was query number 12).
For the subscale Emotion (12 queries): in the No Pattern condition eight queries were significant (exceptions were queries number 7 and 10); in the Partial condition nine queries were significant (all except query number 10), and in the Reveal condition all queries were significant.
For the subscale Imagination (10 queries): in the No Pattern condition eight queries were significant (all except queries number 5 and 10); in the Partial condition nine queries were significant (all except query number 9); for the Reveal condition nine queries were significant (all except query number 3).
For the subscale Intuition (10 queries): in both the No Pattern and in the Partial condition nine queries were significant (except for query number 9 and number 1, respectively); for the Reveal condition all queries were significant.
In Table 1 we listed the significances at 10% significance level [18]. For a cut-off, we introduce an 80% threshold for the whole scale or subscale to be considered of impact in pattern perception.

18 Discussion

In this present study we investigated pattern recognition by humans with a focus on the type of error in connection with religious and supernatural beliefs and thinking styles [7, 8].
The stimuli provided the possibility of studying not only false positive errors (pareidolia) but also false negative errors. To our knowledge, this is the first time that the false negative errors have been taken into consideration, together with false positive errors, when studying a non-clinical population in the context of apophenia in visual perception. We referred to this false negative category of errors as apoidolia (as opposed to pareidolia).
We emphasize that the study presented here is an example of a multidisciplinary approach: where psychological investigation techniques were combined with approaches based on technology and on Bayesian statistics in order to provide novel insights. The stimuli used in previous investigations were often limited to face pareidolia, while we studied general pattern perception that is unrelated to faces. In one study [19], in which the authors used the Rorschach's ink blots as stimuli, complexity was analyzed based on fractals. That approach focusses on describing complexity of non-objects (unclear shapes) triggering the cognitive response.
Our approach focusses on providing the raters with stimuli that are specifically geometrically defined shapes (such as triangles, squares, and other regular polygons) that can be observed in nature.
The main problem was to provide randomness in some rigorous way. We used a random number generator that is state-of-the-art (MATHEMATICA v12.2 from WOLFRAM Technologies®), thereby ensuring a high level of control over the stimulus formation. The use of geometric shapes covered with differing levels of transparency of a random pattern provides absolute control over the pattern presence. Geometric shapes are present in nature in crystals, leaves, nests, blooms, bones, ornaments (butterfly wings, zebra or leopard skin, peacock tail, to name a few), shells, etc. Therefore, the choice of the shapes is relevant for perception processes.
In addition, we present, for the first time, a rigorous Bayesian analysis of differences between Likert-type response sequences (vectors); to our knowledge, this approach of calculation of confusion matrices based on Dirichlet distributions has never been done before.
Among the questionnaires used in the study, the RCI-10 was least able to detect differences between the participants that correctly identified the stimuli in contrast to those who made at least one error. This finding could be due to the generally low religiosity in the specific study cohort, since the respondents were recruited from a scientifically oriented community in a country with a very low religiosity (Czech Republic). In contrast, the best results discriminating between the two groups, independent of the conditions, were obtained using the REIm with each of the subscales having a very high fraction of queries (80% or higher) being significant at 10% level for all the conditions. The Coincidence questionnaire was notably good at distinguishing participants with pareidolia; it performed even better for apoidolia — but only in the Partial condition. We found that, for the IBI, the two subscales Magical beliefs and Spirituality were extremely good at distinguishing subjects with apoidolic perception, specifically in the Reveal condition, but not with pareidolic perception. Interestingly, none of the questionnaires used were better at discriminating between pareidolia and apoidolia. This finding suggests that the presence of an error in the pattern recognition is more important than the specific type of error.
In summary, the results of the current study indicate that, in the context of pattern recognition, the thinking style is more important than the specific belief spectrum of the participants. If the study of visual apophenia involved biological stimuli, such as faces or animal appearances, further cognitive processes may become involved and the role of specific beliefs may become more influential. By using geometrical figures, we avoided such situations and consequently the role of thinking styles emerged in a more identifiable, significant manner.
One limitation of the current study is that our results do not allow us to understand how the overall distribution of responses of the two groups of participants differentiate, even if the results are very reliable in terms of what characteristic is different. To further elucidate this issue, future work is necessary: it will then be necessary to recruit a larger, multicultural sample and incorporate the use of artificial neural networks.

Author Contributions: JB and SB designed the research framework. HP produced the stimuli and chose the frame sequences needed for subsequent analyses. HP designed the statistical tests and wrote the computer programs. SB, HP and JB co-authored the manuscript. All authors contributed to revisions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The research and data collection was supported by Charles University Science Foundation (Grant No. 1404120: "Cognitive Bias, Agency Detection and Pattern Recognition: An Intercultural Quantitative and Experimental Study of Supernatural Belief").

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Faculty of Science, Charles University, Prague, Czech Republic (protocol code 2018/08 approved on April 2, 2018).

Informed Consent Statement: All participants were of age, and provided an informed consent (online) prior to the beginning of the research. All participants were informed that they may revoke their consent and/or terminate the participation at any moment without needing to provide a reason.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Willard, A. K., & Norenzayan, A.: Cognitive biases explain religious be-lief, paranormal belief, and belief in life's purpose. Cognition, 129(2), 379–391 (2013).
2. Kirkpatrick, L. A.: Toward an evolutionary psychology of religion and personality. Journal of personality, 67(6), 921–952 (1999).
3. Jordan, B.: Pattern recognition in human evolution and why it matters for ethnography, anthropology, and society. In Advancing Ethnography in Corporate Environments (pp. 193–214). Routledge (2016).
4. Haselton, M., & Galperin, A.: Error management and the evolution of cognitive bias. Soc. Think. Interpers. Behav. 45, 63 (2012).
5. Fyfe, S., Williams, C., Mason, O. J., & Pickup, G. J.: Apophenia, theory of mind and schizotypy: perceiving meaning and intentionality in randomness. Cortex, 44(10), 1316–1325 (2008).
6. Liu, J., Li, J., Feng, L., Li, L., Tian, J., & Lee, K.: Seeing Jesus in toast: neural and behavioral correlates of face pareidolia. Cortex, 53, 60–77 (2014).
7. Zhou, L. F., & Meng, M.: Do you see the "face"? Individual differences in face pareidolia. Journal of Pacific Rim Psychology, 14, e2 (2020).
8. Van Leeuwen, N., & Van Elk, M.: Seeking the supernatural: The interactive religious experience model. Religion, Brain & Behavior, 9(3), 221–251 (2019).
9. Weinberger, A. B., Gallagher, N. M., Warren, Z. J., English, G. A., Moghaddam, F. M., & Green, A. E.: Implicit pattern learning predicts individual differences in belief in God in the United States and Afghanistan. Nature communications, 11(1), 1–12 (2020).
10. Wardle, S. G., Taubert, J., Teichmann, L., & Baker, C. I.: Rapid and dynamic processing of face pareidolia in the human brain. Nature communications, 11(1), 1–14 (2020).
11. Pascalis, O., & Kelly, D. J.: The origins of face processing in humans: Phylogeny and ontogeny. Perspectives on Psychological Science, 4(2), 200–209 (2009).
12. Mandelbrot, B. B., & Mandelbrot, B. B.: The fractal geometry of nature (Vol. 1). New York: WH freeman (1982).
13. Bressan, P.: The connection between random sequences, everyday coincidences, and belief in the paranormal. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 16(1), 17–34 (2002).
14. Worthington Jr, E. L., Wade, N. G., Hight, T. L., Ripley, J. S., McCullough, M. E., Berry, J. W., ... & O'Connor, L.: The Religious Commitment Inventory-10: Development, refinement, and validation of a brief scale for research and counseling. Journal of counseling psychology, 50(1), 84 (2003).
15. Kingdon, Bianca L., Sarah J. Egan, and Clare S. Rees.: The Illusory Beliefs Inventory: A new measure of magical thinking and its relationship with obsessive compulsive disorder. Behavioural and Cognitive Psychotherapy 40.1, 39–53 (2012).
16. Norris, P., & Epstein, S.: An experiential thinking style: Its facets and relations with objective and subjective criterion measures. Journal of personality, 79(5), 1043–1080 (2011).
17. Jackson, D. L.: Review of The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century, by David Salsburg. Structural Equation Modeling: A Multidisciplinary Journal, 10(4), 651–655 (2003).
18. Caelen, O.: A Bayesian interpretation of the confusion matrix. Annals of Mathematics and Artificial Intelligence, 81(3), 429–450 (2017).

19. Taylor, R. P., Martin, T. P., Montgomery, R. D., Smith, J. H., Micolich, A. P., Boydston, C., ... & Spehar, B.: Seeing shapes in seemingly random spatial patterns: Fractal analysis of Rorschach inkblots. PloS one, 12(2) (2017).

Appendix

Consider Fig. A-1. We observe two heat maps of a questionnaire and a stimulus (Coincidence questionnaire for the stimulus in Reveal condition): one heat map for the numbers of entries for 'Correctly identified' and one for 'At least once wrong'. Visually, the heat maps appear different. We ask whether the distribution of entries for each query are *significantly* different. The significance can be determined by calculating the confusion matrix — one for each query. In Fig. A-1, the confusion matrices for each query are displayed in the graph.

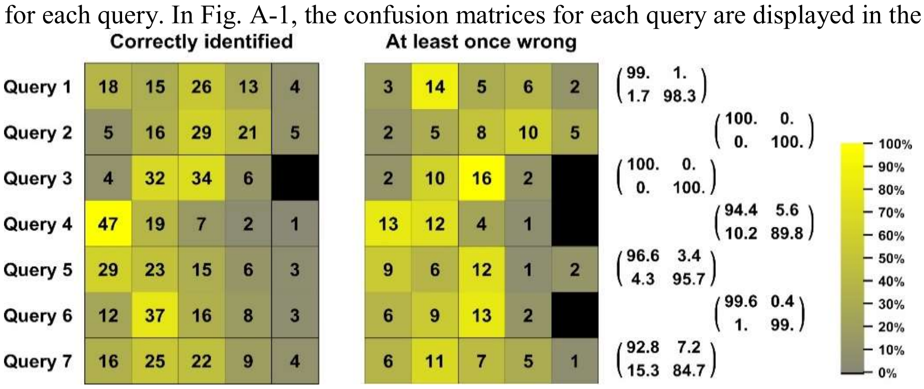


Fig.A-9. The heat map of one questionnaire with the entries and the confusion matrices displayed. Observe that the sum of the number of entries for each query for 'correctly identified' is not the same as for 'at least once wrong'. Black squares occur when there are no entries. The shade of yellow, on the other hand, identifies the fraction of occurrences. Thus, for example, the 3rd entry for Query 5 for 'Correctly identified' is darker than the same entry for 'At least once wrong', although numerically the entries are comparably close.

We simplify the notation: one heat map we label A and the other B; the query we label q. To calculate the confusion matrix with entries $\begin{pmatrix} p_{A_{\text{TRUE}}} & p_{A_{\text{FALSE}}} \\ p_{B_{\text{FALSE}}} & p_{B_{\text{TRUE}}} \end{pmatrix}$, where $p_{A_{\text{TRUE}}}$, etc., are the probabilities of the respective entries. We cannot calculate these entries using the Frequentist (Laplace) approach for two reasons: (1) the entries are too small to be considered a satisfactory limit of large sample size, and (2) we do not know the (analytic) boundary between TRUE and FALSE. We therefore use Bayesian statistics. We first describe the method for a one-dimensional problem (Fig. A-2), in which we can define a boundary. Consider a query with only two entries for 'Correctly identified' (A) ($\{s_{A_1}, 1-s_{A_1}\} = \{\,37 \quad 12\,\}$) and only two entries for 'At least once wrong' (B) ($\{s_{B_1}, 1-s_{B_1}\} = \{\,6 \quad 9\,\}$). For each query, the entries are Beta-distributed (Fig. A-2), with likelihood functions $\mathcal{L}_A(s)$ and $\mathcal{L}_B(s)$.

In the one-dimensional problem (Fig. A-2), calculating the confusion matrix is straightforward, because the probabilities of the likelihoods are integrals. Beyond the one-dimensional distributions of entries for a query, however, the problem is very much more difficult, because there is no boundary that can be analytically specified for integration: the likelihood functions are Dirichlet distributions. In these cases, we use Monte Carlo methods to find the probabilities.

We note that for a randomly chosen value for s_A in Fig. A-2, the condition TRUE is when $\mathcal{L}_A(s_A) > \mathcal{L}_B(s_A)$; likewise for a randomly chosen value for s_B in Fig. A-2, the condition TRUE is when $\mathcal{L}_B(s_B) > \mathcal{L}_A(s_B)$. This determination of TRUE and FALSE can be used in higher dimensions as well, because $s_A = \{s_{A_1}, \cdots, s_{A_k}\}$ (for k entries for a query) and the (pseudo) random number generator is defined for the Dirichlet distribution $\mathcal{Dir}\left(s_{A_1}+1, \ \cdots, s_{A_k}+1\right)$, irrespective of the number k of different entries for a given query. If n_{ran} random numbers are generated (in the paper presented here, $n_{ran} = 175000$), then the confusion matrix is $\begin{pmatrix} \frac{n_{A_{\text{TRUE}}}}{n_{ran}} & \frac{n_{A_{\text{FALSE}}}}{n_{ran}} \\ \frac{n_{B_{\text{FALSE}}}}{n_{ran}} & \frac{n_{B_{\text{TRUE}}}}{n_{ran}} \end{pmatrix}$. These ratios in the confusion matrix are not the probabilities defined by the Frequentist (Laplace) limit, but rather the Monte Carlo method of approximating an integral — equivalent to the Bayesian probability as graphed in Fig. A-2 in the one-dimensional case.

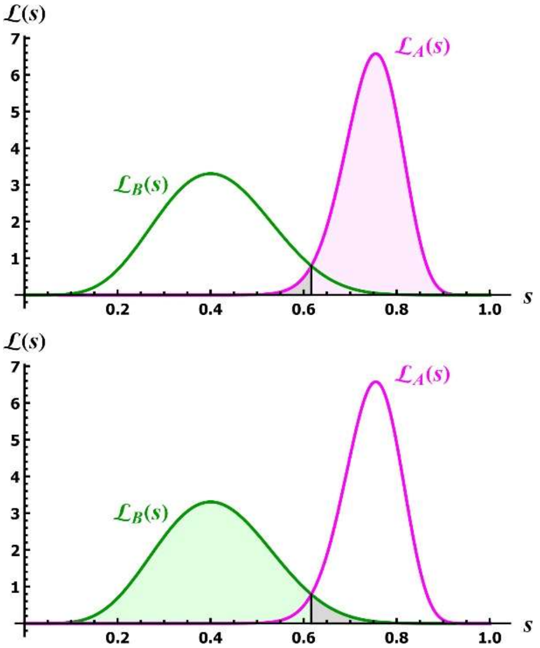


Fig. A-2. The likelihood functions for two samples (A, magenta) ($\{s_{A_1}, 1 - s_{A_1}\} = \{37 \quad 12\}$) and ($B$, green) ($\{s_{B_1}, 1 - s_{B_1}\} = \{6 \quad 9\}$). The likelihood functions are Beta distributions. In the top graph, the probability of $\mathcal{L}_A(s)$ for TRUE is shaded light magenta, while the probability for FALSE is shaded gray. In the bottom graph, the probability of $\mathcal{L}_B(s)$ for TRUE is shaded light green, while the probability for FALSE is shaded gray. The vertical black line shows the boundary between TRUE and FALSE. In this (numerical) example, the gray area in the top graph is 0.0235, and in the bottom graph the gray area is 0.0437; the confusion matrix is therefore $\begin{pmatrix} 0.976 & 0.0235 \\ 0.0437 & 0.956 \end{pmatrix}$. The distributions are therefore significantly different at 10% significance level [12].

Quantifying the Rating Performance of Ambiguous and Unambiguous Facial Expression Perceptions Under Conditions of Stress by Using Wearable Sensors

Jakub Binter[1,2](✉), Silvia Boschetti[1,2], Tomáš Hladký[1,2], Hermann Prossinger[3], Timothy Jason Wells[1], Jiřina Jílková[1], and Daniel Říha[1,2]

[1] Faculty of Social and Economic Studies, University of Jan Evangelista Purkyně, Ústí nad Labem, Czech Republic
jakub.binter@ujep.cz
[2] Faculty of Humanities, Charles University, Prague, Czech Republic
[3] Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria

Abstract. *Background*: In real-world scenarios humans perceive the world contextually, relying on previous information to modify their responses. During interactions with a machine, missing contexts may decrease the accuracy of judgements. In the realm of human-computer interactions (HCI), relatively easy tasks as controls may not be relevant.

To evaluate the impact of stress we increased the cortisol level by the safe but reliable procedure Cold Pressor Task. We used five stimuli represented by facial expressions: 'neutral', 'laughter', 'fear', 'pain', and 'pleasure'.

Aim: We intend to find out how the responses to stimuli are altered by stress and statistically quantify the BVP (Blood Volume Pulse) signals.

Materials: 27 raters rated these five stimuli presented by 5 actors and 5 actresses, while BVP was being registered.

Methods: Each physiological response was a six-second time series after the rater rated the stimulus. A nontrivial model includes lag dependencies on either previous states or previous noise. The simplest models would be $ARMA(p, q)$ models with to-be-determined parameters $\varphi_1, \ldots \varphi_p$ and $\theta_1, \ldots \theta_q$.

Inferences: In this study, we find that the wearables' sampling for six seconds cannot separate signal from noise significantly. Only one response was found to be significantly affected by the condition of stress: the perception of fear.

Keywords: ARMA time series · 'Wearables' · BVP · Maximum likelihood distributions · Bayesian likelihood · Emotion perception · Fear · Pain · Pleasure · Stress

1 Introduction

In real-world scenarios, humans estimate a large proportion of their perceived world contextually and use previous information to adjust or modify their expectations and

M. Kurosu et al. (Eds.): HCII 2022, LNCS 13516, pp. 519–529, 2022.
https://doi.org/10.1007/978-3-031-17615-9_36

responses. During interactions with a machine, contexts may be missing and, therefore, it is claimed, the accuracy of judgements decreases [6].

Furthermore, many applications using human-machine-interfaces rely on the premise that ratings and responses are in concordance and, consequently, requesting a response for a rating would not be necessary, since the signal can be directly obtained from the sensors (installed, for example, in wearables) [2, 7].

However, in the realm of human-computer interactions (HCI), using relatively simple tasks as controls may not be relevant; in the real world, ambiguous situations and stress presumably negatively impact the measurable response.

To evaluate the impact of stress we have employed a procedure commonly used in psychological experiments called Cold Pressor Task (see below). It reliably, but safely, increases the cortisol level. The participants were not informed that stress will be induced (so as to eliminate any possible bias) [4].

To evaluate the impact of possible ambiguity we used five stimuli present in human faces: two of which are easily identified by the raters ('neutral' and the basic emotion 'laughter'), one basic emotion that is often misinterpreted ('fear'), as well two affective states ('pain' and 'pleasure').

One of the main reasons for using the sensors for the communication between human and machine is to create a short-cut and allow for multiple channels to detect the state of the human responder or, alternatively, to categorize the response directly without explicitly asking the human via interviews or questionnaires [9].

There are certain tasks that are automatic for humans, many of which are visual-based e.g., the gaze oriented towards moving target. [5] As stimuli, the use of facial expressions is an ideal task since this is a process that is automatic and triggers a response. Indeed, allocation of attention towards face or face-like stimuli is one of the automatic, fast processes that develops from early childhood onwards [9].

In this study we aimed to test whether the condition of increased arousal affects physiological responses, namely Blood Volume Pulse (BVP), to the displayed stimuli.

Previous articles have shown that for similar tasks the stationary measuring devices (sampling rate of 256 Hz) are necessary for at least 5 s of exposition to the stimulus. [5] Usually multiple sensor data are required (BVP, EDA, temperature) but the wearables we used had a presumably insufficient sampling rate other than for BVP.

2 Materials and Methods

2.1 Data Collection

This sample consists of 27 students ($N_{♂} = 14$ and $N_{♀} = 13$) aged 19–30 years. Some of them ($N_{\text{stress}} = 15$) underwent a procedure called Cold Pressor Task which consists of immersing a subject's limb in ice water (2–4 °C) for 90 s. This procedure causes an elevation of the stress hormone cortisol causing an increase of the sympathetic nervous system activation. The sympathetic nervous system is functionally related with the psychological concept of arousal. [4] Consequently, the participant responded under a physiological state of high arousal (stress). The control group underwent the same procedure but the water was of room temperature so there was no effect on physiological arousal.

The participants were wearing E4 Empatica® (MIT) bracelets to collect data about their physiological reaction during the procedure. In this study, we focus on the output from the photoplethysmography (PPG) sensor with a sampling frequency of 64 Hz, namely the Blood Volume Pulse (BVP). BVP is derived from a PPG process, which uses the infrared light reflected by the skin to estimate blood vessel diameter. Diameter changes of the peripheral blood vessels, which are regulated by the sympathetic neural system, affect the amount of light reflected back to the photo-sensor; this changes the amplitude of the signal, corresponding to sympathetic activation. Figure 1 shows ten signals for one stimulus for one individual.

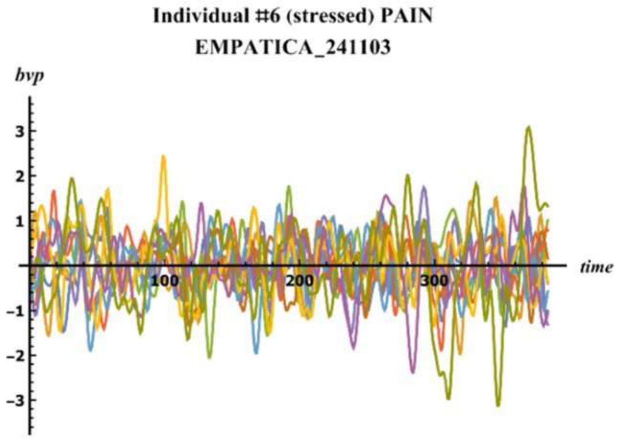


Fig. 1. Ten time-series of a stressed rater while assessing the stimulus ‘pain’ exhibited by five actors and five actresses. Different time series are rendered with a different color. Unit of time is $\frac{1}{64}$ s. The first measurement is ‘immediately’ (i.e. within 1 ms) after the rater has pressed the enter button that registers his/her rating of the stimulus; the last measurement is 384 pulses later. (Color figure online)

The procedure consisted of presentation of visual stimuli. In total 50 stimuli were displayed: five types of facial expressions representing emotions and affective states (laugh, pain, pleasure, and pain), along with neutral. Each facial expression was displayed by five actors and five actresses and each participant rated them on a scale A–C. For this study, we only analysed the BPV signal, not the ratings.

2.2 Statistical Analyses

Some of the time series had extraordinary positive and negative values. These, furthermore, did not meet the criterion of weak stationarity. They were discarded. We also tested whether the (rare) discarded time series were uniformly distributed across all stimuli and whether the raters were control or stressed. We used the Bayesian approach to test for significant differences. If n_1 is the number of discarded time series for control raters (for a given stimulus) and n_2 the number of discarded time series for the stressed raters, then the Bayesian likelihood of the probability s that the number of discards is greater for the controls is $\Lambda(s) = \frac{\Gamma(n_1+n_2+2)}{\Gamma(n_1+1)\Gamma(n_2+1)} s^{n_1}(1-s)^{n_2}$ where $\Gamma(\cdot)$ is the Gamma function. [1] If the mode is close to $s = \frac{1}{2}$, and the $\mathrm{HDI}_{95\%}$ uncertainty interval [3] overlaps the probability $s = \frac{1}{2}$, then there is no significant difference between the discarded time series for control versus stressed.

We use this Bayesian method of computing the likelihood function $\Lambda(s)$ [1] and determining the uncertainty interval $\mathrm{HDI}_{95\%}$ [3] for numerous other comparisons below.

Distribution of peaks: we subtracted a Laplacian filter of each time series from the time series to identify the peaks. First, we simply inventoried the number of peaks in each signal for each stress. We estimated the ML distribution for each signal from a suite of three distributions: lognormal, Weibull and Gamma. We then used the log-likelihood of the ML distribution of peaks for control and stressed and Wilks lambda to find the probability that the control and stressed peaks were derived from the same statistical population. Formally: if $\ln\Lambda_C$ is the log-likelihood for the control, and $\ln\Lambda_S$ is the log-likelihood of the stressed, then Wilks lambda $\Lambda_{\mathrm{Wilks}} = -2(\ln\Lambda_{CS} - (\ln\Lambda_S + \ln\Lambda_C))$. is χ^2-distributed with $\mathrm{df} = \mathrm{df}_S + \mathrm{df}_C - \mathrm{df}_{CS}$ degrees of freedom. (The index CS refers to the data set formed by combining the data set for C with that for S.)

We constructed a matrix: the (ten) rows were the ratings by a participant of a specific stimulus and in each row were the 384 sequential registrations by the wearable. We first tested whether smoothing via Singular Value Decomposition (SVD) could be used to eliminate noise. None of the scree plots of the squares of the singular values showed a knee. In fact, the square of the first singular value never explained more than 20% of the square of the Frobenius norm of the matrix.

We therefore used time series analysis to investigate the signal for each rating by each individual for each stimulus—a total of $27_{raters} \times \left(5_{stimuli} \times (5_{♀} + 5_{♂})\right) = 1350$ time series, minus the 32 discarded ones. We used an ARMA (*p*,*q*) time series model for each time series. In an ARMA(*p*,*q*) model, the signal is modeled as $y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$ with both $p > 0$ and $q > 0$. The functions w_{t-k} are Gaussian white noise contributions with finite variance $\sigma_w^2 > 0$. [8] The p coefficients (amplitudes) $\varphi_1 \cdots \varphi_p$ describe the contributions of the p past values of the underlying signal and the q coefficients (amplitudes) $\theta_1 \cdots \theta_q$ describe the previous/lagged white noise contributions to the observed signal; c is a constant. We use a software package (MATHEMATICA v12.4 from WOLFRAM Technology) to find a fit for the maximum p and q (and their numerical estimators) using AICc (Akaike's Information Criterion, corrected for finite sample size) as the optimality criterion. We also tested for weak stationarity.

Since we are interested in whether the time series of the raters depended on whether they were controls or were stressed, we tallied all numbers of amplitudes p for each stimulus, both for the stressed and for control. Because the largest number (by far) was for $p = 3$ (see below), using a χ^2-test for homogeneity is not meaningful, because we will find the probability of homogeneity to be very low. Instead, we use a Bayesian approach by comparing, for each $p = 0 \ldots p = 6$ for each stimulus, the frequencies of the stressed and the controls. We then use the machinery of computing the likelihood function $\Lambda(s)$ [1] and determining the uncertainty interval $\mathrm{HDI}_{95\%}$ [3].

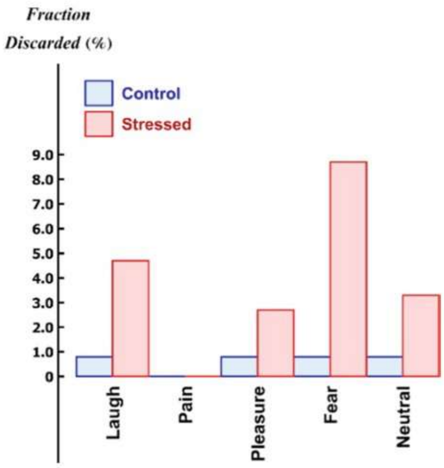


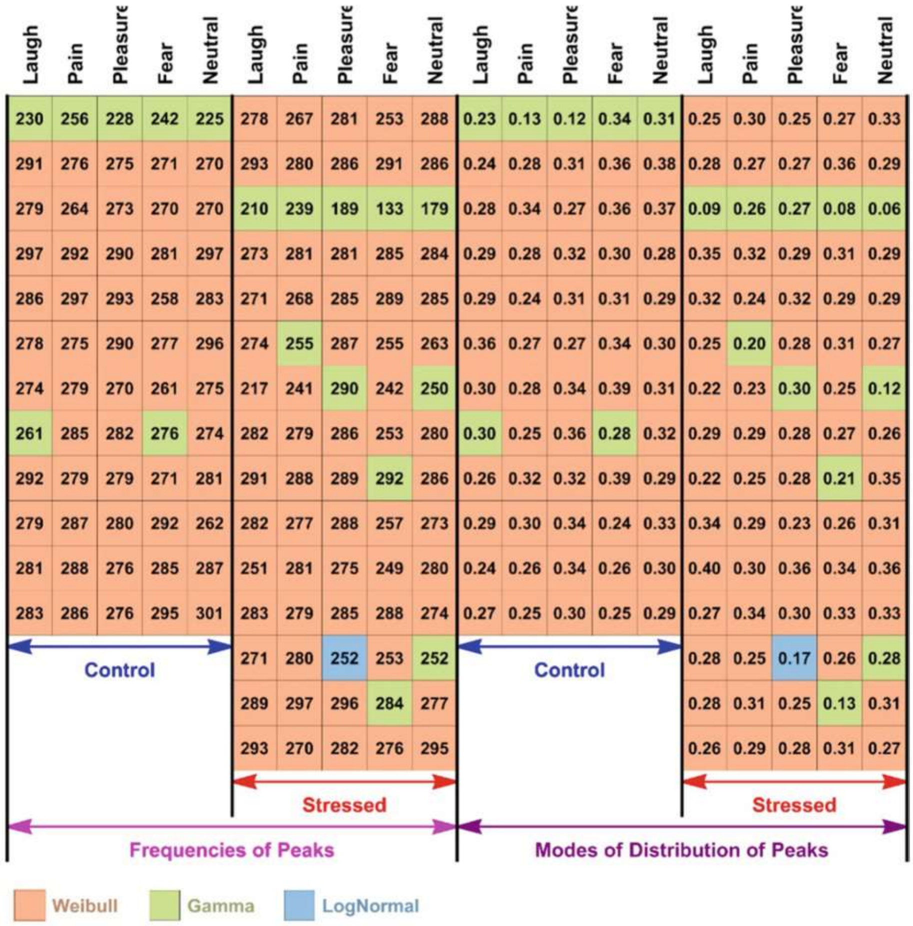
Fig. 2. A bar chart of fractions of time series that had to be discarded, prior to statistical analyses, because of extremely large departures from the predominantly occurring amplitudes. We note that these departures occurred more often when the raters were stressed, but the fractions were still small. No time series needed to be discarded for the stimulus 'Pain'. A Bayesian analysis of the fractions 'control' versus those 'stressed' showed a significant difference only for 'Laugh' and 'Fear' (at 5% significance level).

Table 1. The results of the Wilks lambda test of whether the peaks for one stimulus versus that of another stimulus is significantly (at 5% significance) different: control (light blue background) versus stressed (light orange background) raters. The percentages are the fractions of the compared stresses that were significantly different.

	Laugh	Pain	Pleasure	Fear	Neutral
Laugh		16.7%	33.3%	33.3%	41.7%
Pain	33.3%		33.3%	16.7%	25.0%
Pleasure	33.3%	26.7%		41.7%	8.3%
Fear	46.7%	40.0%	40.0%		33.3%
Neutral	33.3%	40.0%	40.0%	40.0%	

3 Results

A bar chart of the 32 discarded time series is shown in Fig. 2. Table 1 shows the fractions of pairwise significant differences in peak distributions of stimuli. A tests for significance (using Wilks lambda) of numbers of peaks control versus stress show that only one pair ('Fear' control versus 'Fear' stress) is significantly different (results not shown). Likewise, a test for a significant difference for intervals/gaps between peaks showed no significant difference (results not shown). Per individual, the ML distributions of numbers of peaks as well as the modes of their distributions are most often Weibull distributed, a few Gamma distributed and only one log-normally distributed (Fig. 3).



Frequencies of Peaks: Control					Frequencies of Peaks: Stressed					Modes of Distribution of Peaks: Control					Modes of Distribution of Peaks: Stressed				
Laugh	Pain	Pleasure	Fear	Neutral	Laugh	Pain	Pleasure	Fear	Neutral	Laugh	Pain	Pleasure	Fear	Neutral	Laugh	Pain	Pleasure	Fear	Neutral
230	256	228	242	225	278	267	281	253	288	0.23	0.13	0.12	0.34	0.31	0.25	0.30	0.25	0.27	0.33
291	276	275	271	270	293	280	286	291	286	0.24	0.28	0.31	0.36	0.38	0.28	0.27	0.27	0.36	0.29
279	264	273	270	270	210	239	189	133	179	0.28	0.34	0.27	0.36	0.37	0.09	0.26	0.27	0.08	0.06
297	292	290	281	297	273	281	281	285	284	0.29	0.28	0.32	0.30	0.28	0.35	0.32	0.29	0.31	0.29
286	297	293	258	283	271	268	285	289	285	0.29	0.24	0.31	0.31	0.29	0.32	0.24	0.32	0.29	0.29
278	275	290	277	296	274	255	287	255	263	0.36	0.27	0.27	0.34	0.30	0.25	0.20	0.28	0.31	0.27
274	279	270	261	275	217	241	290	242	250	0.30	0.28	0.34	0.39	0.31	0.22	0.23	0.30	0.25	0.12
261	285	282	276	274	282	279	286	253	280	0.30	0.25	0.36	0.28	0.32	0.29	0.29	0.28	0.27	0.26
292	279	279	271	281	291	288	289	292	286	0.26	0.32	0.32	0.39	0.29	0.22	0.25	0.28	0.21	0.35
279	287	280	292	262	282	277	288	257	273	0.29	0.30	0.34	0.24	0.33	0.34	0.29	0.23	0.26	0.31
281	288	276	285	287	251	281	275	249	280	0.24	0.26	0.34	0.26	0.30	0.40	0.30	0.36	0.34	0.36
283	286	276	295	301	283	279	285	288	274	0.27	0.25	0.30	0.25	0.29	0.27	0.34	0.30	0.33	0.33
					271	280	252	253	252						0.28	0.25	0.17	0.26	0.28
					289	297	296	284	277						0.28	0.31	0.25	0.13	0.31
					293	270	282	276	295						0.26	0.29	0.28	0.31	0.27

Fig. 3. The frequencies of peaks and modes for each individual for each stimulus, along with the ML distributions of these peaks. The largest number of peaks is Weibull distributed, very few are Gamma distributed, and only one is log-normally distributed. Most modes are Weibull distributed, very few are Gamma distributed, and only one is log-normally distributed. No modes are greater than 1.

Very rarely was the observed signal dependent on more than three previous signals (as opposed to the noise component of the signal; Fig. 4). For all stimuli, the amplitudes φ_p ($p \leq 3$) are collinear and do not significantly differ between controls and stressed (Fig. 5). For every p (the number of nonzero contributions φ_p to the AR part of the ARMA(p,q) time series), there was no significant difference between control and stressed individuals (Fig. 6).

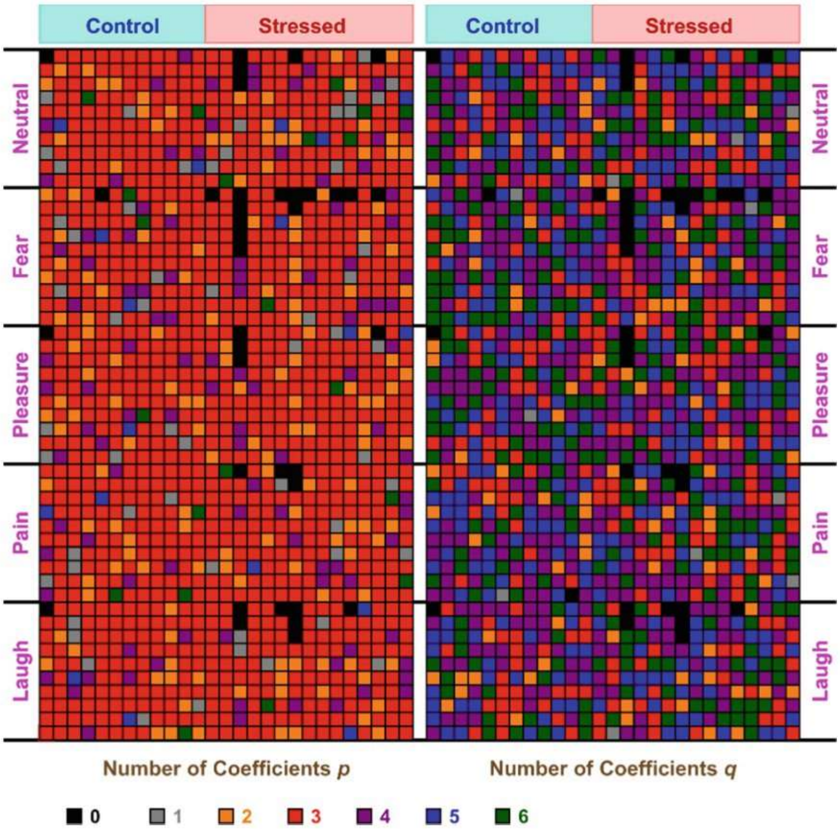


Fig. 4. The heat map of the frequencies of the coefficients φ_k $k = 1 \ldots p$ and θ_j $j = 1 \ldots q$ in the ARMA(p,q) time series models for the signals supplied by the wearables' data sets. The numbers of coefficients for θ are randomly distributed, whereas those for φ are not. In fact, by far the most $\varphi_k k = 1 \ldots p$ are for $p = 3$. There is no significant statistical difference between the frequency distribution of coefficients for θ between 'control' and 'stressed'; nor is there any for the coefficients for φ. In this latter case, very, very many are $p = 3$.

These results show that the BVP signal for each individual provides no evidence that inducing stress altered the distribution of peaks nor the distribution of the amplitudes φ_k in the ARMA(p,q) time series, except for some features of 'Fear'.

4 Discussion

The data set we analyze here is large: there are 1350 time series with 384 data points each. The measurement routine can be considered reliable, as only 32 of 1350 (2.4%) time series had to be discarded. It is remarkable that the non-noise amplitudes φ_k in the ARMA(p,q) time series are co-linear (Fig. 5). We cannot explain this and have found no reference to such a relation in the literature. This linearity implies, for example, that the amplitudes φ_1 and φ_2 are anti-correlated with each other and anti-correlated with φ_3. In other words, if a BVP signal (without noise) is strongly dependent on the previous amplitude, then the dependence on the 'pre-previous' amplitude is strongly suppressed, and by almost the same amount. Whether this effect depends on latencies in the blood vessel dilation, we do not know (yet).

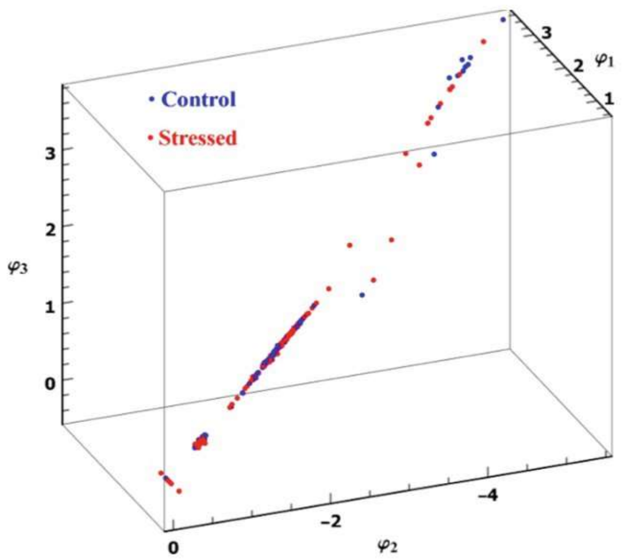


Fig. 5. A graph of the amplitudes φ_k $k = 1 \ldots 3$ for all ARMA(3,q) time-series registered by the wearables as the raters rated the stimuli showing 'Pain'—color-coded according to whether the raters were control or stressed. Not all time-series are rendered in this graph, because some have been discarded, and some have ARMA(p,q) models with $p > 3$. Displayed are 110 'control' time-series and 146 'stressed' time-series. In this graph, for all φ_k $k = 1 \ldots p < 3$ we observe that $\varphi_k = 0 \, \forall \, 2 \leq k < 3$. These (few) φ_k can be seen as a short row of dots at $\varphi_3 = 0$. (Color figure online)

The primary goal of the study was to investigate whether the wearables can detect an alteration in the time series, after the cortisol level is raised. We find no significant effect. One possible explanation for this may be that 6.00 s (384 measurements) is too short for the circulatory system to respond. Another possibility is that effects are only detectable in a multivariate setting: we have not investigated a vector time series. In any case, the univariate signal BVP provides no evidence for a significant difference between stimuli for an individual (control versus stressed). Outside a laboratory setting, the decision to decrease the sampling frequency is probably motivated by extending battery life and maintaining state tracking, which lasts longer than stimulus exposure.

The outcome regarding the distribution of the ML distributions is important for the development of algorithms that will be applied in future computations. We expect most ML distributions to be Weibull.

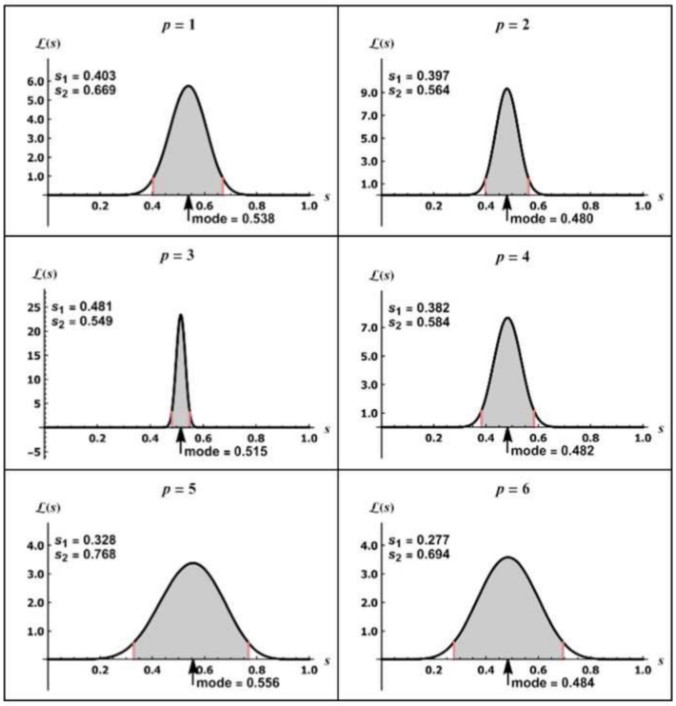


Fig. 6. The result of testing whether the amplitudes φ_k $k = 1 \ldots p$ of the ARMA(p,q) are significantly different for the controls versus the stressed. The probability s is the probability that amplitudes for a given p are more probable for the 'controls'. For all six p, the $HDI_{95\%}$ uncertainty interval (s_1, s_2) overlaps the probability $s = \frac{1}{2}$, so there is no evidence of a significant difference between the amplitudes φ_k $k = 1 \ldots p$ for 'stressed' versus 'controls'. All modes are very close to $s = \frac{1}{2}$. The short, red, vertical lines show the borders of the $HDI_{95\%}$ interval. (Color figure online)

We point out that in one case a significant difference was found, namely in case of 'Fear'. This is an important finding involving several research fields. In the field of psycho-physiological research, there is an ongoing debate about the problem of fear perception regarding categorization (positive vs. negative), and classification (labeling the stimulus with a correct name). It is present in many sensory fields (visual-facial perception, and acoustic-vocalization perception). Our study may resolve this problem by providing support for involvement of an inner state (in our case stress) in ambiguous stimuli perceptions.

5 Conclusion

This paper shows that there are shortcomings when using wearables related to monitoring stimuli data. Although their use is on the increase, it should be pointed out that, currently, wearables are not precise laboratory devices; they have limitations due to short battery life, sensitivity to movement artifacts (accelerations), and data gathering frequencies incompatible with several physiological variations (such as blood vessel dilation). In our case, the sampling rate only allowed us to use one sensor output, namely BVP.

It should be pointed out though that the proportion of discarded data is below 3%, which is a very good result, indicating that the wearables could handle almost any challenges posed by the wearer.

Usually, wearables are used in applications that last several minutes or more. While we argue that the six-second period is insufficient, we have no conclusive evidence other than the hint of a very small signal-to-noise ratio (Fig. 4). We point out that this novel approach (using time series analyses) is one rigorous method of quantifying noise. We have not found literature analyzing noise in wearables applications with such short sampling times. The oftentimes used analysis deals with amplitudes of peaks; we point out that perhaps too many (noise-related) artefacts are then included in the analyses. If wearables are to be used as assistance to human-computer-interaction, the sampling rate needs to increase, with sufficient battery-life, the wearables all the while remaining sufficiently unobtrusive for the wearer. It should be further noted that the future use of human-computer-interaction should also include rapidly changing signals, e.g., EEG, combined with several signals from wearables so as to generate a multivariate time series. Thus, high-sampling rate wearables could supply valuable input about the individual´s psychological state, such as stress-level, which will inform researchers about the central-nervous system response measures. There is an urgent need to develop the methods in the near future since in many situations the state of the individual may be a life-threatening one.

Acknowledgements. The research and data collection was supported by Czech Science Foundation (Grant No. 19-12885Y "Behavioral and Psycho-Physiological Response on Ambivalent Visual and Auditory Stimuli Presentation"). Part of the research team (JB, SB, TH, JJ) is employed by the Project SMART reg. no. CZ.02.1.01/0.0/0.0/17_048/0007435 Smart City-Smart Region-Smart Community. D.Ř. is funded by the Ministry of Education, Youth and Sports, Czech Republic and the Institutional Support for Long-term Development of Research Organizations, Faculty of Humanities, Charles University, Czech Republic (Grant COOPERATIO "Arts and Culture").

Conflict of Interest. The Author(s) Declare no Competing Interests.

Ethics Statement. The Project Was Evaluated and Approved by the Ethical Committee of the Faculty of Science, Charles University, as Part of Broader Project (7/2018). GDPR Regulations Were Followed at All Times.

References

1. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)

2. Gao, R., Islam, A., Gedeon, T., Hossain, M.: Identifying real and posed smiles from observers' galvanic skin response and blood volume pulse. In: Yang, H., Pasupa, K., Leung, A.C.-S., Kwok, J.T., Chan, J.H., King, I. (eds.) ICONIP 2020. LNCS, vol. 12532, pp. 375–386. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-63830-6_32
3. Kruschke, J.K.: Doing Bayesian Data Analysis. A Tutorial with R, JAGS, and STAN. Elsevier, San Diego (2015)
4. Lamotte, G., Boes, C.J., Low, P.A., Coon, E.A.: The expanding role of the cold pressor test: a brief history. Clin. Auton. Res. **31**(2), 153–155 (2021). https://doi.org/10.1007/s10286-021-00796-4
5. Maaoui, C., Pruski, A.: Emotion recognition through physiological signals for human-machine communication. Cutting Edge Rob. **11**, 317–332 (2010)
6. Santos, L.R., Rosati, A.G.: The evolutionary roots of human decision making. Annu. Rev. Psychol. **66**, 321 (2015)
7. Sindhu, N., Jerritta, S., Anjali, R.: Emotion driven mood enhancing multimedia recommendation system using physiological signal. In: IOP Conference Series: Materials Science and Engineering, vol. 1070, no. 1, p. 012070. IOP Publishing, February 2021
8. Shumway, R.H., Stoffer, D.S.: Time Series Analysis and its Applications. With R Examples. Springer, New York (2006). https://doi.org/10.1007/978-3-319-52452-8
9. Tcherkassof, A., Dupré, D.: The emotion-facial expression link: evidence from human and automatic expression recognition. Psychol. Res. **85**(8), 2954–2969 (2021)

Visual Analysis of Emotions Using AI Image-Processing Software: Possible Male/Female Differences between the Emotion Pairs "Neutral"–"Fear" and "Pleasure"–"Pain"

H. Prossinger
Department of Evolutionary Biology, Faculty of Life Sciences, University of Vienna, Vienna, Austria
hermann.prossinger@univie.ac.at

T. Hladký
Faculty of Humanities, Charles University, Prague, Czech Republic,Faculty of Science, Charles University, Prague, Czech Republic
hladkytomas@seznam.cz

J. Binter
Faculty of Humanities, Charles University, Prague, Czech Republic,Faculty of Science, Charles University, Prague, Czech Republic
jakub.binter@natur.cuni.cz

S. Boschetti
Faculty of Humanities, Charles University, Prague, Czech Republic,Faculty of Science, Charles University, Prague, Czech Republic
boschetti.silvia@gmail.com

D. Říha
Faculty of Humanities, Charles University, Prague, Czech Republic,Faculty of Science, Charles University, Prague, Czech Republic
daniel.riha@fhs.cuni.cz

ABSTRACT

Inferring the emotional state of an individual by viewing his/her facial expression seems to be present in all human cultures. Numerous studies have shown that various changes in facial muscles determine the resulting facial expression. The analysis of images of faces expressing emotional states promises to contribute to quantification of the claimed observations. Here, we use a suite of AI (artificial intelligence) algorithms, along with ML (maximum likelihood) estimated distributions to quantify the shift in facial expression from "neutral"→"fear" and "pain"→"pleasure". The images are single frames of five emotional states (neutral, fear, pain, pleasure, laugh) expressed by actors and actresses in BDSM videos. We extract a feature vector for each image, dimension-reduce these feature vectors by mapping them onto a two-dimensional manifold and calculate the norms of the normalized displacement vectors for each emotional pair. We then find that the ML distributions of the norms are Gamma-distributed and that the modes for each pair are different for both males and females. We use Wilks lambda to determine significance. We find that the distributions for the females are significantly different, but not for the males. The methodology we present here has widespread applications: monitoring the emotional states of humans in various settings; among these: determining whether participants in BDSM and similar videos are indeed volunteering their participation or are victims of criminal activity.

PETRA 2021, June 29–July 02, 2021, Corfu, Greece

ACM ISBN 978-1-4503-8792-7/21/06…$15.00
https://doi.org/10.1145/3453892.3461656

CCS CONCEPTS

• **Affective Computing**; • **Behavior Monitoring Systems**; • **Computer Vision in Healthcare**; • **Human-Robot Interaction**; • **Smart Cities of the Future**; • **Technologies for Improving the Quality of Daily Living**; • **Vocational Safety and Health Monitoring**;

KEYWORDS

Image Processing using AI (artificial intelligence), Maximum likelihood methods for Gamma distributions, Facial expression of emotions, Facial expression of fear, of pain and of pleasure

ACM Reference Format:
H. Prossinger, T. Hladký, J. Binter, S. Boschetti, and D. Říha. 2021. Visual Analysis of Emotions Using AI Image-Processing Software: Possible Male/Female Differences between the Emotion Pairs "Neutral"–"Fear" and "Pleasure"–"Pain". In *The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021), June 29–July 02, 2021, Corfu, Greece.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3453892.3461656

1 INTRODUCTION

Interpersonal interaction is often thought to be conveyed by verbal exchanges, albeit not exclusively. Facial expressions are considered a further natural method to "advertise" one's emotional affective states during the majority of social interactions. This latter claim is supported, it is argued, by their universal perception [3, 9].

The ability to express emotional affective states and the further ability to ascertain the displayed emotional affective state has been shown to be universal. A considerable body of intercultural studies have provided evidence that the particular emotional expressions joy, fear, sadness, and anger are to a large degree universally understood; the same seems to hold for the expression of pain (which is not a basic emotion) [3, 6].

Some emotional expressions are so closely linked to emotional states that there exists a consensus among psychologists that a small number of pure emotions can be considered basic states, much like base vectors in a coordinate decomposition of a point in some high-dimensional space. These base states are characterized by specific — and identifiable — facial muscle contractions [6]. They are identifiable not only by researchers active in psychological field work, but also by anyone observing these facial expressions in someone else's face.

This has led to codification of facial muscle movements called FACS (Facial Action Coding System) which is based on so-called Action Units. Each Action Unit is a collection of facial muscle contractions (and releases) that create, through an as yet-to-be-determined synchronization, a particular facial expression [6]. The method is used descriptively (in order to analyze the emotional status of an individual) but also prescriptively as a template for the animation of facial expressions [11]. Animations are routinely used in the computer graphics industry, where the distinction between the facial expressions of pain and of sexual climax is particularly interesting; furthermore, a difference of mental representations of these two emotional expressions could be derived [5].

One can, of course, down-regulate the expressiveness and thereby affect the amount of information shared between the experiencing person and his/her observer (the proverbial *poker face* comes to mind) but there may be limits that apply to extreme experiences. In the context of sexuality, this grimace issue has already been described in the "Kinsey Reports" [7]. Recently, other examples from other social situations have also been observed [12].

Pain, pleasure and fear are three emotional states suspected to be ambiguous to observers/raters whenever these states are not presented within the encompassing context of the communication. Interestingly, the accuracy of judgement of these emotional states can be even lower when actors/actresses are instructed to act the emotion rather than when it is expressed in genuine, real-life settings [1]. Thus, BDSM (Bondage-Dominance-Sadistic-Masochistic) sexual acts, performed by professional actors in commercial for-sale videos are a data set we can use to investigate the expressions of these emotions and pain.

Use of BDSM videos of private persons would infringe on their right to privacy and we could not reliably infer which emotion is being expressed (we would have to rely on their statements). In the case of BDSM videos with professionals, on the other hand, we are certain that the displayed emotion is intentionally expressed, perhaps not genuinely. If the intensity is insufficiently displayed, the video, along with the actor/actress, would disappear from the market. We cannot expect the emotions expressed by the actor/actress to be translated into a genuine grimace, but this uncertainty is the leverage point we use for our analysis.

Here, we pursue an image analysis regimen, including AI, to focus on high-intensity states during which the facial muscles are contracted to a grimace, putatively communicating an otherwise difficult-to-identify intensive experience. We quantify the differences in grimaces by using the Euclidean distance between normalized dimension-reduced feature vectors, extracted from the images of the grimaces, to derive the statistics of differences and their significance.

We pursue this analysis methodology because we are not certain whether the emotion pair transitions "neutral"→"fear" and "pain"→"pleasure" are quantifiable and whether they are sufficiently useful for a large range of image analysis applications.

2 MATERIALS

One of the authors (T.H.) scanned ten BDSM videos with professional participants, five of females and five of males. Based on the development of the plot in the video, five frames (one frame of neutrality, one of fear, one of pleasure, one of pain, one with laughter) were chosen from each video. The data base thus consists of 50 images of these five emotions expressed by five actors and five actresses.

3 METHODS

We could not expect the emotions expressed by the actor/actress in a pornographic video to be grimaces that express genuinely felt emotions, but the pornographic industry and its marketing mechanisms ensure that the grimaces in successfully sold videos must be convincingly genuine for the viewer. We therefore used such videos in this investigation.

We (a) used AI image recognition software to extract the rectangle containing the face with negligible borders in each of the 50 frames (one for each of the five emotions expressed by the 10 actors/actresses); (b) used a suite of image analysis routines supplied by the software package [13] to align the five faces for each actor/actress in each video; (c) applied a feature extraction function to the five images thus aligned to generate four 4D feature vectors (one for "neutral", one for "pleasure", one for "fear", and one for "pain") for each actor/actress; (d) dimension-reduced each of these 4D feature vectors to coordinates on a 2D manifold via an auto-encoder neural network with 7 layers; and (d) calculated the Euclidean distance between (normalized) pairs of these dimension-reduced 2D feature vectors for each video (Fig. 1 & 2).

We then estimated the ML distribution of these Euclidean distances and calculated the modes as well as the $HDI_{95\%}$ uncertainty for the distance distribution between the emotion pairs "neutral"→"fear" and "pain"→"pleasure". (The $HDI_{95\%}$ interval is a confidence interval; it is the generalization of the $\pm 1.96s$ interval for the normal distribution, where in this formula s is the point estimator of the standard deviation of the sample. The $HDI_{95\%}$ is derived from the ML distribution, not from the point estimators of the sample.) We then used Wilks lambda to determine whether the ML distributions (and, by implication, their modes) for the emotion pairs "neutral"→"fear" and "pain"→"pleasure" are significantly different.

4 RESULTS

For the females, the ML distributions for the emotion pairs "neutral"→"fear" and "pain"→"pleasure" are Gamma distributions with modes 1.177 and 0.791, respectively (Fig. 3a); Wilks lambda is 5.985, so the probability that both emotional distances are drawn from the same statistical population is 5.0%. In other words, the two populations of (normed) distances are significantly different. We also observe that the mode of the normed distance "neutral"→"fear" is greater than the mode of the normed distance "pain"→"pleasure". We observe that the $HDI_{95\%}$ uncertainty for the females during the transition "neutral"→"fear" is not only least of all four (males and females), but remarkably narrow.

For the males, the ML distributions for the emotion pairs "neutral"→"fear" and "pain"→"pleasure" are Gamma distributions

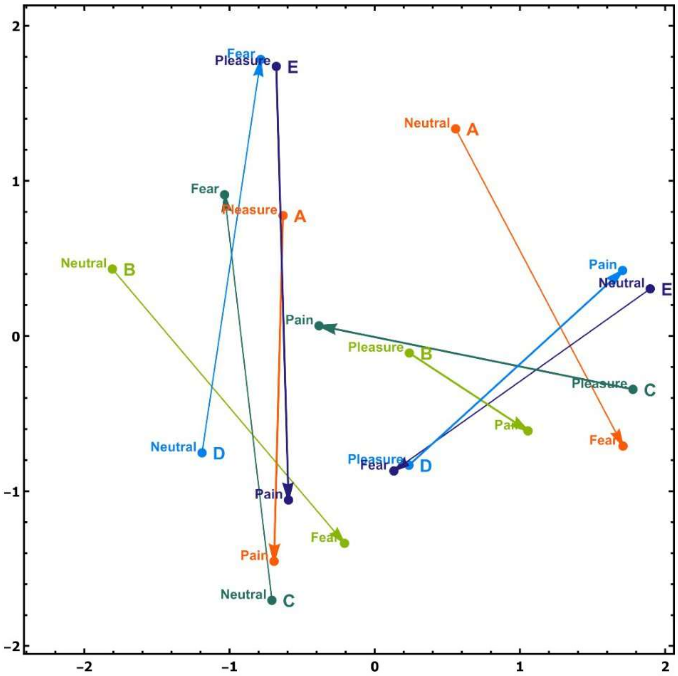


Figure 1: The shifts in emotion manifest in the images of the five females, obtained by dimension-reduction of the feature vectors of the image frames. Color-coding is unique for each female; arrows show the direction we investigated in this study; capital letters identify the females. There are two arrows per female, because we investigated two shifts. We observe that the shifts "neutral"→"fear" are oftentimes longer than "pain"→"pleasure".

with modes 0.801 and 1.158, respectively (Fig. 3b); Wilks lambda is 3.658, so the probability that both distances are drawn from the same statistical population is 16.1%. In other words, the two populations of (normed) distances are insignificantly different (at 5% significance level). We also observe that the mode of the normed distance "neutral"→"fear" is less than the mode of the normed distance "pain"→"pleasure" — opposite the mode shifts for the females. However, because the different distributions are not significantly different, this inequality is of no consequence.

5 DISCUSSION

We looked for the five leading ML distributions (from more than a dozen, including mixture distributions) in order to choose one. We used three criteria for our choices: (a) the domain of the distribution must be positive (thus, the normal distribution must be excluded), (b) AIC (Akaike's Information Criterion) must be a minimum, and (c) if there are competing distributions meeting criteria (a) and (b), the ones that agree for both the emotion pairs "neutral"→"fear" and "pain"→"pleasure" (but not necessarily the same distribution family for males and females). Using these three criteria, we find that all the emotion distances are Gamma distributed (Fig. 3).

To assume that the normed distances are Student *t*-distributed is fallacious, so a *t*-test is not possible. Furthermore, because ML methods estimate both ML parameters of the distribution in one sweep, an ANOVA test is doubly useless (ANOVA assumes that the sample points are drawn from a normal or nearly-normal distribution and that the variance estimated from the data is a good estimator of the variance of the distribution).

Each stimulus/image used in this study was one single frame taken from the video at the moment an intensive affective state was shown/experienced by the actor/actress; we relied on the researcher making the choice, based on his following the plot while watching the video (or relevant parts of it). The resolution of the extracted image was occasionally as low as 500×550 pixels (because the face was within a BDSM scene); furthermore the orientation of the head was not predetermined. We used an image alignment algorithm [13] to ensure reasonable alignment of the faces via image rotation. The feature extraction software is minimally dependent on the

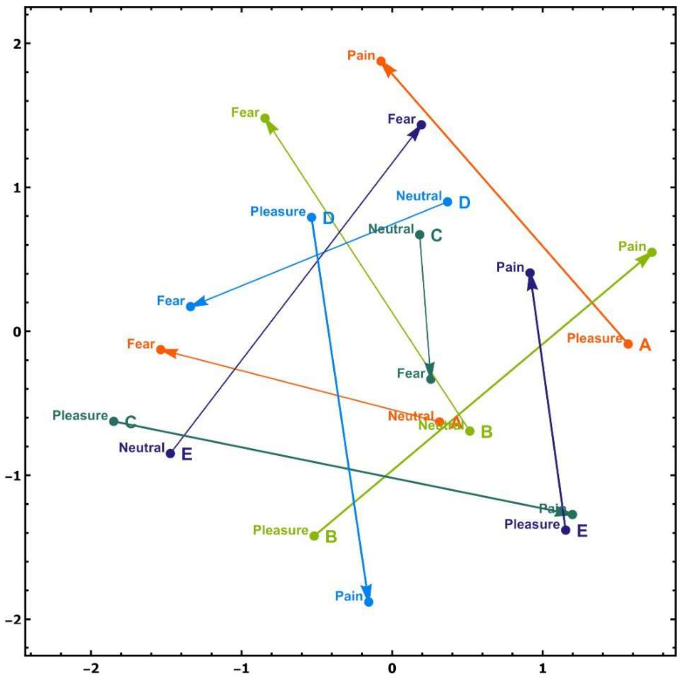


Figure 2: The shifts in emotion manifest in the images of the five males, obtained by dimension-reduction of the feature vectors of the image frames. Color-coding is unique for each male; arrows show the direction we investigated in this study; capital letters identify the males. There are two arrows per male, because we investigate two shifts. We observe that the shifts "neutral"→"fear" are not always longer than those for "pain"→"pleasure".

exactitude of the alignment, so we could (and did) assume the results are neither biased by the alignment issue nor the smallness of the image of the face (if the image is not too small). Consequently, the result is remarkable in terms of accuracy, far more so than a result obtainable by some human rater. Evidence: we have previously applied the same method of feature vector extraction of images in short horror-video game clips [9] where 50 frames (with chosen maximum expression in the 24th frame) were used to successfully distinguish neutral facial expression from expressions of joy and fear, respectively.

6 CONCLUSION

The differentiation of distinct facial expressions can be further used in the implementation of safety fuses during personal use of horror videos at home as suggested by [2] but also for maintaining the well-being of smart-city inhabitants (for a review see [8]).

The method we presented here can potentially be used in the analysis of already existing video materials (typically online) to determine whether the depicted individuals are personally endangered or are experiencing discomforting — or worse — situations [2]. Needless to mention, it is important to determine the border between BDSM videos that have voluntarily participating actors/actresses and those that inflict criminal violence. The proof-of-principle we demonstrate here can contribute to keeping the market for BDSM videos legal. The method definitely has its limitations if and when the quality of the data set is insufficient.

This novel approach we have presented here demonstrates that the emotional pairs pain and pleasure are easily distinguishable from each other in males and females with surprisingly high accuracy — likewise for the pairs consisting of neutral facial expression and fear. The outcomes bring novel approaches to emotion differentiation applied to semi-naturally occurring videos (typically depicting the interactions of two adult individuals participating in BDSM activities). Other novel approaches to be developed are the analysis of the highly salient expression of ambiguous emotions by employing further AI algorithms available in MATHEMATICA® [13].

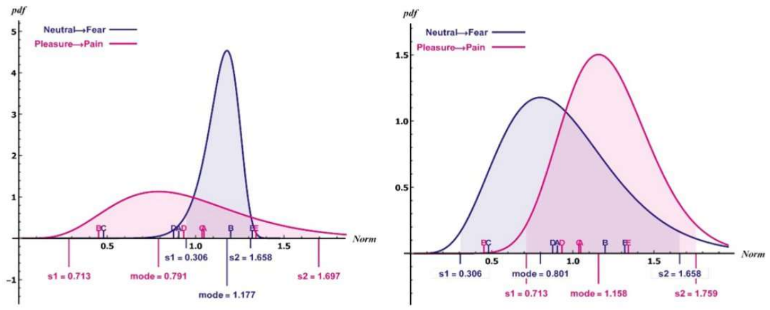


Figure 3: The ML distributions of the shifts "pain"→"pleasure" (dark magenta) and "neutral"→"fear" (dark blue) for the normalized Euclidean distances (females: Fig. 1 and males: Fig. 2) for the females (above left) and the males (above right). The color-coded rug plots in both graphs show the normalized Euclidean distances. All ML distributions are Gamma distributions; the shaded areas (from s1 to s2, respectively) are 0.95 for each distribution; these define the $HDI_{95\%}$ confidence intervals. We observe that the modes of all distributions differ and for both males and females the $HDI_{95\%}$ confidence intervals overlap. However, the two sets of normalized distances for the female shifts are significantly different (at 5% significance level), while for the males they are not.

These methods generate numerous implications for emotion research, namely in health care, in on-the-spot diagnostics (such as when a potential witness is being interviewed by a detective), and in applications in the fields of information and communication technologies. The method can be used to not only estimate one's affective state with high accuracy during certain activities and to trigger safety fuses [2] but also to retrospectively evaluate existing visual materials. Importantly, since video-recording is a non-invasive type of data collection and real-time analysis is available, we stress that it should become a method of choice to benefit people.

ACKNOWLEDGMENTS

TH, JB, SB & DŘ are funded by the Czech Science Foundation (Grant No. 19-12885Y "Behavioral and Psycho-Physiological Response on Ambivalent Visual and Auditory Stimuli Presentation"). The stimuli were obtained from freely available online videos from webpages that offered direct possibility of download. The current study is part of broader set of projects that were approved by Ethics Committee of Faculty of Science, Charles University (reg. no. 2018/08).

REFERENCES

[1] Abramson, L., Marom, I., Petranker, R., & Aviezer, H. (2017). Is fear in your head? A comparison of instructed and real-life expressions of emotion in the face and body. Emotion, 17(3), 557.
[2] Binter, J., Říha, D., & Prossinger, H. (2020, July). A Warning: Potential Damages Induced by Playing XR Games. In International Conference on Human-Computer Interaction (pp. 260–270). Springer, Cham.
[3] Ekman, P. (1993). Facial expression and emotion. American Psychologist, 48(4), 384–392.
[4] Ekman, P. (2002). Facial action coding system (FACS). A human face.
[5] Chen, C., Crivelli, C., Garrod, O. G., Schyns, P. G., Fernández-Dols, J. M., & Jack, R. E. (2018). Distinct facial expressions represent pain and pleasure across cultures. Proceedings of the National Academy of Sciences, 115(43), E10013–E10021.
[6] Gendron, M., & Barrett, L. F. (2009). Reconstructing the past: a century of ideas about emotion in psychology. Emotion Review, 1(4), 316–339. doi:10.1177/1754073909338877
[7] Kinsey, A.; Pomeroy, W.; Martin, C., & Gebhard, P. Sexual Behavior in the Human Female, Philadelphia: Saunders (1953), ISBN 978-0-253-33411-4.
[8] Muhammad, G., Alsulaiman, M., Amin, S. U., Ghoneim, A., & Alhamid, M. F. (2017). A facial-expression monitoring system for improved healthcare in smart cities. IEEE Access, 5, 10871–10881.
[9] Prossinger, H., Binter, J., Hladky, T., & Říha, D. (2021) Using neural-network-driven image recognition software to detect emotional reactions in the face of a commentator while playing a horror video game. HCI in Games, Conference Proceedings ID 3143.
[10] Valente, D., Theurel, A., & Gentaz, E. (2018). The role of visual experience in the production of emotional facial expressions by blind people: A review. Psychonomic bulletin & review, 25(2), 483–497.
[11] van der Struijk, S., Huang, H. H., Mirzaei, M. S., & Nishida, T. (2018, November). FACSvatar: An Open Source Modular Framework for Real-Time FACS based Facial Animation. In Proceedings of the 18th International Conference on Intelligent Virtual Agents (pp. 159–164).
[12] Wenzler, S., Levine, S., van Dick, R., Oertel-Knöchel, V., & Aviezer, H. (2016). Beyond pleasure and pain: Facial expression ambiguity in adults and children during intense situations. Emotion, 16(6), 807.
[13] Wolfram Research, Inc. Mathematica, Version 12.2, Urbana-Champaign, IL (2020).

Nature Scientific reports WOW and AHH

"Wow!" and "Aah": Bimodal Stimuli of Affective States with High and Low Intensity Displays

Jakub Binter[1], Hermann Prossinger[1,3], Tomáš Hladký[1] and Silvia Boschetti[1,2]*

[1] Faculty of Humanities, Charles University, Prague, Czech Republic

[2] Faculty of Science, Charles University, Prague, Czech Republic

[3] Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria

***Correspondence to:**
Silvia Boschetti, Faculty of Humanities, Pátkova 2137/5, 182 00 Praha, Czech Republic
silvia.boschetti@natur.cuni.cz
ORCID: 0000-0002-8048-4062

Keywords: Pain, Pleasure, Facial Expressions, Vocalization, Due-to-Chance, Dirichlet Distribution, Confusion Matrix, Sensory Dominance
Running title: Bimodal stimuli of high and low affective states

Abstract

Our research consists of studying the probability of humans being able to assess the differences between the valence of human faces in combination with simultaneous human vocalizations of high (pain and pleasure) and low (smile/laughter and neutral expression/speech) intensities. The study was conducted online and used a large sample ($n = 902$) of respondents. The task was to categorize whether human vocalizations and facial expressions that can be considered semi-naturalistic were rated positive, neutral, or negative when presented with audio stimuli and pictures of faces. These had been extracted from freely downloadable online videos. Each rating participant (rater) was presented with four facial expressions (stimuli), accompanied by the simultaneous vocalizations. Two of these were highly intense (pain and pleasure) and two of low intensity (laugh/smile and neutral). Using a Bayesian statistical approach, we could test for consistencies and due-to-chance probabilities. The outcomes support the prediction that the results (ratings) are not due to chance in all cases (so some ratings were not guesses, even though they might have been incorrect) — findings agreeing with result from unimodal auditory rating but not in agreement with unimodal facial expressions ratings. The highly intense displays are incorrectly attributed. Therefore, we can assume that the auditory information is dominant in terms of certainty of the rating; yet it does not provide extra information for the case of highly intense affective expressions.

Introduction

When interacting, in real life with others (an ecologically valid scenario) and interpreting their affective states, we humans are exposed simultaneously to a dynamic multitude of information about the expresser (i.e. facial expression, body posture, vocalization, etc.) and, in many cases, about the situation eliciting such states (i.e. context, antecedents, etc.) [1, 2, 3, 4, 5, 6, 7].

The concerto of information that is processed via different channels (often auditory or visual channels; olfaction is yet to be evaluated but suggestions exist — e.g., [8]). In such a situation, it is hard to assess the importance that single elements (i.e. facial expression or vocalization) have in determining the interpretation of the affective states.

The study of the ability to correctly attribute emotional and affective expressions of others was impacted by recent papers showing counterintuitive evidence that affective states with high intensity are more difficult to correctly assess from facial expression [9, 10, 11, 12] or vocalization [13, 14] without further context.

There are two main characteristics of affective states and emotion valence and intensity that have been studied. The approach postulates that the two vectors of all effects be placed somewhere in the plane. The same rules apply for all modalities; this two-dimensional model of emotion [15, 16] can also be used when physiological measures or self-report measures are implemented, making it an ideal tool. In special cases, we can limit one the analysis to one vector — for example, when testing extremely intensive affect displays or displays specifically pre-selected to address this issue [9, 11, 12].

This approach has been used in previous studies in which the visual (facial expression) and auditory (vocalization) stimuli were presented separately to investigate the ability to correctly identify the emotional valence (positive/negative) of the stimulus while being presented with only one modality. The published literature on the topic suggests that the facial expression of affective states was correctly attributed with high accuracy when their intensity was low (i.e. laugh and neutral expression) but they were not correctly assessed when the intensity was high (i.e. pain and pleasure). This effect was found also for emotional vocalization, leading to the conceptualization of the "emotional intensity paradox" (a terminology first so named by Holz and colleagues [14]; even though earlier publications on the topic exist): vocalization emitted in high arousal effective states are harder to correctly categorize as a positive or negative emotion [14, 17].

Other studies showed, however, that further information provided to the observer can increase the correct identification of the highly intensive emotion. For example, the facial expressions when winning or losing are not recognized alone but they are correctly categorized when the accompanying body posture was also shown. This may be an effect of some specific source of information: body posture is better recognized than the face in the specific case of winning or losing [9], therefore the clarity of one source of cues (the body) may disambiguate the less clear source of the other cue (facial expression). Alternatively, the better rating could be explained by the extra information presented to the observer and another type of information source (i.e. vocalization) could bring about the same effect.

The most recent studies show that there are further important measures — other than correctness — that should be taken into account. Two studies [12, 17] suggested (each for a different modality) that the results need to be consistent, and not due to chance (equivalent to guessing).

The use of Bayesian statistical analysis by the above-mentioned researchers shows that the respondents were highly consistent (they reliably rated the stimuli twice in differently randomized orders) and not guessing (see the description in the Appendix) when rating the affective states of low intensity (neutral and laugh in both modalities).— distributed almost equally between the positive and negative rating options. The ratings were due-to-chance (the possibility of not guessing was under 1%; [12]) for the visual modality. An interesting comparison is provided by the results of auditory stimuli ratings – the outcome is equally unsuccessfully assessed but there but the ratings were not due to chance — the chance of guessing was under 1% [17].

It should be pointed out that the stimuli used in the last two aforementioned studies are semi-naturalistic and were not produced in a laboratory since the laboratory approach was highly criticized for its low ecological validity [12, 17, 18]. The used stimuli also were not pre-tested (a step that is often used by researchers to evaluate the stimuli being in concordance with the mental representation of the population). Furthermore, this pre-testing is impossible for extremely intensive facial and vocal displays since they are inherently ambiguous [9,11, 12, 17].

Previous studies of two modalities show an additive effect when they are congruent and competitive when incongruent [19, 20, 21, 22, 23, 24, 25, 26].

Therefore, it is highly interesting to investigate whether the congruent combination of visual or auditory stimuli will increase the accuracy of judgment (an additive effect) and whether the due-to-chance probability will be shifted towards one of the senses (sensory dominance).

The Current Study

Our study is designed to overcome some of the aforementioned methodological limitations.

Aim: to quantify the consistency of the ratings of affective facial expressions and vocalizations, focusing on the (biological) sex of the rater as well as the (biological) sex of the rated (the expressers). The stimuli used were images of facial expressions and audio records of four emotional vocalizations, with high (pain and pleasure) and low (neutral and smile/laugh) intensity. The facial expression and the vocalization were presented simultaneously, but randomized. As in previous studies, the stimuli were produced by different male and female expressers [27] to avoid biases due to the expressivity of the expresser.

We generated our set of stimuli by using picture frames and audio records from videos that depict consensual acts of extreme sexual activities. The stimuli were previously used in three studies: Prossinger [28] used facial expressions to evaluate the differences between facial expressions of pain and pleasure using AI (Artificial Intelligence) based methods; it showed there is a difference between these facial expressions. The same set of stimuli was used by Boschetti [12] to further support the claim that human raters are unable to correctly categorize facial expressions. The vocalizations during the same time points as the picture frames were used by Binter [17] to evaluate the ability of raters to assess the valence of vocalizations; they obtained similar accuracy of ratings outcomes.

We retained the categorical methodology; there are three rating categories: positive, negative, or neutral. We predicted that, if the individual is exposed to a negative stimulus (pain, for instance) — the grimace on the picture and the vocalization on the audio record (with the expected high intensity) — will be rated as negative. In a manifestly opposite stimulus (pleasure, for instance) the rating should be positive.

The general expectation is that female raters perform better than male raters when rating negative emotions [29, 30, 31, 32]. In one study, female raters performed better in recognizing female expressions of pain [10]. However, female raters were not more successful in rating the valence of stimuli in the previous two studies using unimodal highly intensive affect stimuli (facial expression: [12]; vocalization: [17]).

We also expect that ambiguous stimuli of high intensity presented as bimodal (though audio and visual channels) will have a higher probability of correct identification in comparison with the same stimuli presented in a unimodal presentation (only audio or only visual channel).

Materials & Methods

Raters and Stimuli

Expressers' vocalizations and Expressers' Facial Expressions: In order remain consistent with the published terminology, we use the terms "expressers", "faces", "vocalizations", and "vocal displays" to describe what was presented in the video frames and audio records as stimuli. We specify the biological sex of each expresser (evident from the videos) with male and female.

Raters: In order to remain consistent with the published terminology, we use the terms "expression raters" and "respondents" to describe the individuals who were presented with the stimuli and who provided their ratings. We specify the biological sex of the expresser with the terms male and female. The biological sex we list is the respondent's self-reported one. We deleted all ratings ($n = 4$) who did not report their biological sex.

A total of 902 individuals (aged 18–50; M_{age} = 32 years, SD = 8.9 years) completed the questionnaires; 526 women (M_{age} = 30.9 years, SD = 8.3 years) and 376 men (M_{age} = 33.6 years, SD = 9.5 years). 61,08% of the participant has secondary education or lower and 61.75% live in an area with less than 100000 people.

Criteria for inclusion were: (a) age of respondents between 18 and 50 years, and (b) at least a minimal experience with adult media, since the vocalizations used in this study were extracted from such materials.

The data were collected in the Czech Republic in 2021 via the agency Czech National Panel (narodnipanel.cz) and a science-oriented online portal pokusnikralici.cz using the online platform for data collection Qualtrics™. Participants submitted responses either via computer keyboard or touchscreens of mobile devices (smartphones or tablets).

The stimuli (4 affective states — laugh, neutral, pleasure, and pain — expressed by female and male expressers) were presented to the subject in random order for 1.5 seconds and rated as positive, negative, or neutral.

Stimuli generation

From the numerous audio-visual materials ten were chosen (five with female expressers and five with male expressers). Based on the plot in each of these ten, five frames with faces and simultaneous vocalizations were selected (one each of neutrality, fear, pleasure, pain, and smile/laughter). Three of the authors (S.B. J.B. & T.H.) are researchers in field of human sexuality with more than ten years of experience of focusing on extreme sexual behavior and on consumption of erotic materials. All three authors (one female and two male) provided their opinion on all of the chosen stimuli. All agreed on stimuli choice and what expression is to be expected, based on the contextual information. The agreement on stimuli choice was debated among all three researchers in dedicated meetings.

We point out that a common misconception is that the individuals taking part in such exchanges derive sexual pleasures from pain and the two (pain and pleasure) happen simultaneously. Although this may be possible, we have found no mention for such evidence in the published scientific literature. Rather, it should be noted that sensitivity is increased by the experience of pain in various parts of the body (in our case mainly slapping the buttocks and thighs) and only after the painful procedure is over is climax achieved. All the chosen visual and audio stimuli were derived from the context of the video. Specifically, we could rely on the images/scenes to identify the displayed emotions and affects. There is no doubt, due to the camera perspective, about the occurrence of the climax in male expressers. No such explicit judgment can be used for female expressers, but all signals of the occurrence of climax were identified by the researchers (involving breathing, contraction of pelvic musculature, twitching of anal sphincter muscles, facial blushing, vocalization, etc.; [33]), and further supported by expressers' self-reports at the ends of the videos.

In each video, male/female faces and vocalizations expressed pain and pleasure during the session, while smile and laughter and neutral facial expressions and vocalization (speech) were filmed and recorded during an interview prior to the pain and pleasure experiences. All picture stimuli presented to the raters were scaled to 600×600 pixels; we used triangulation between the tip of the nose and pupils to ensure that the proportions of the face on the screen were comparable among all frames. No background was visible within the frames presented. All audio stimuli were adjusted to the same sound level and lasted from 0.5 seconds to 1.5 seconds— depending on the stimulus.

Ratings

Previous literature [34] has noted that it is a challenging task to correctly identify the facial expression (e.g., to categorize the expression of pain as pain). The same can be claimed about human vocalizations. We, therefore, asked our participants to rate the observed expression as either positive, neutral, or negative. We thereby avoided the problem of correct labeling and avoided any intricacies associated with a verbal categorization system. The ratings were communicated either via using keyboard keys or a touchpad with dedicated areas (specified by icons); they were subsequently stored in a dedicated database.

Statistical Analyses

Due to the inherent advantages of Bayesian statistics when dealing with our research questions, we implemented this approach. General descriptions follow, while more detailed descriptions, augmented by a graphical display, are provided in the Appendix I.

The statistics we evaluate is the mode of the *pdf* of each rating distribution, which is a 3-parameteric Dirichlet distribution. The components of each mode are the probabilities of rating positive, negative or neutral by the raters of the expressers.

Confusion matrices: The method of determining whether two groups are significantly different (or not) is to calculate the confusion matrix; it is the obligatory method to use when sample sizes are small. One sample (*F*) has a distribution $dist_F$ and another sample (*G*) has a distribution $dist_G$. When there is an overlap of the *pdf*s (probability density functions) of these two distributions, a fraction of *F* is TRUE (and a fraction is FALSE); likewise, for *G*. The confusion matrix has four entries:

$$\begin{pmatrix} \text{TRUE}_F & \text{FALSE}_F \\ \text{FALSE}_G & \text{TRUE}_G \end{pmatrix}$$

If both off-diagonal elements ($\{\text{FALSE}_F,\ \text{FALSE}_G\}$) are small, there exists a significant difference between the distributions of *F* and of *G* (the significance level being chosen by the researcher). Observe that the sum of each row in the confusion matrix is $1 = 100\%$. The fractions in the confusion matrix can also be calculated using Monte Carlo methods.

Possibility of effects being due to chance: In Bayesian statistics, the probability s is a random variable ($0 \le s \le 1$). For each combination of raters and expressers s is the probability of a correct rating. Thus, the likelihood function is the probability density function of a Beta distribution (Appendix 1). The crucial separator for determining chance is $s = \frac{1}{2}$. The probability is either the integral of the likelihood function $\mathcal{L}(s)$ over the interval $0 \le s \le \frac{1}{2}$ or the integral over the interval $\frac{1}{2} \le s \le 1$, depending on which side of $s = \frac{1}{2}$ the mode is. In either case, the integral determines whether an observation is due to chance. (A graphical description is shown in the Appendix I).

Results

Age distributions by biological sex

Table 1 and Fig.1 show that the distributions of the male and female raters have different modes. We use KDE (kernel density estimation) with a Gaussian kernel because, as Fig. 1 shows, it is not to be expected that raters of either sex have a parametric distribution or even a superposition of one or two such parametric distributions. We also note that modes and expectation values differ between the sexes and also for the same sex. The $HDI_{95\%}$ uncertainty interval is very broad, so we can consider the distributions for both sexes to be comparable to a uniform distribution of respective ages. The confusion matrix shows that the two distributions are not significantly different.

Table 1

Estimator	Male Raters	Female Raters
N	376	526
Range (years)	18–50	18–50
Mode (years)	41.7	27.0
$\mathbb{E}$ (years)	33.6	30.9
Mean (years)	33.6	30.9
$HDI_{95\%}$ (years)	16.4–50.3	15.8–49.8

Table 1 The descriptors and the estimators for the ages of the raters, separated by (biological) sex. $\mathbb{E}$ is the expected value (estimated by $\mathbb{E} = \int_0^{60} u \times pdf(\text{KDE}, u)\, du$, where KDE is the kernel density estimation), and $HDI_{95\%}$ (highest density interval: the interval with a 95% probability of observing an age;[35]).

Figure 1

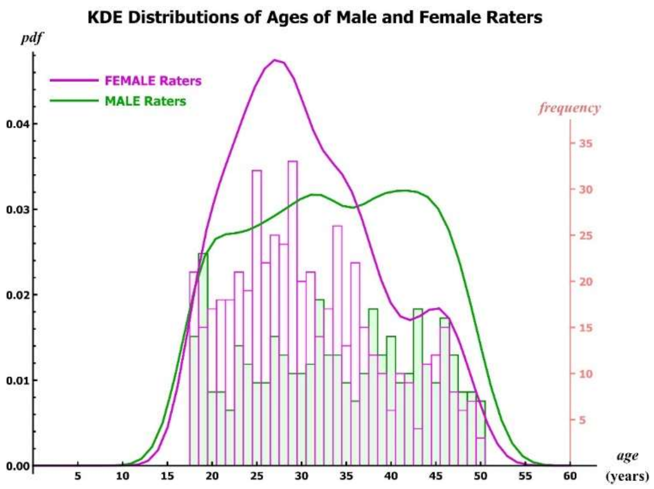


Figure 1 The distributions of the ages of males and of females estimated using a KDE (kernel density estimation) with a Gaussian kernel. Superimposed of the graphs of the *pdf*s is a histogram of the registered ages (scale of frequencies on the right). Modes, expected values, means, and $HDI_{95\%}$ of both distributions are listed in Table 1. The two distributions are not significantly different; the confusion matrix is $\begin{pmatrix} 46.3 & 53.7 \\ 28.3 & 71.7 \end{pmatrix}\%$

Confusion Matrices: Sex Difference in Ratings

We used the confusion matrix to test for significant differences between male and female raters, separately for male and female stimuli. An example of the Beta distribution used for

the calculation of the confusion matrix is in Fig. 2. We assume that the results are significant if both of the off-diagonal entries in the confusion matrix are less than 10% [36]. The Dirichlet distributions of the stimuli rating are in Appendix II.
Figure 2

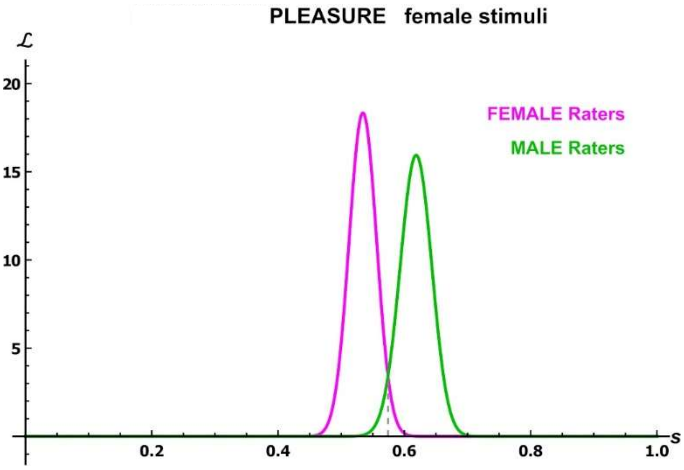


Figure 2 One example of the two Beta distributions for the affective state pleasure for female stimuli — female and male raters. s is the probability of correct rating; $\mathcal{L}$ is the likelihood of the probability s. The thin, dashed, vertical line is the boundary between true/false for each distribution. The integrals to the left and right of this vertical line are used for the calculation of the confusion matrix (numerical values in Table 2). Details are described in the text. The modes of the probabilities (the abscissa values) are the ML (most likely) probabilities of the likelihoods of each distribution; numerical values are in Table 2.

For the stimuli produced by female expressers (Table 2), we observed significant differences between the ratings of female and male raters for the high arousal affective states (pain and pleasure), while we do not observe any significant differences for the low arousal affective states (laugh and neutral). In this last case, the accuracy of the rating was high for both, male ($\text{mode}_{\text{neutral}} = 0.832$; $\text{mode}_{\text{laugh}} = 0.832$) and female ($\text{mode}_{\text{neutral}} = 0.855$; $\text{mode}_{\text{laugh}} = 0.819$) participants. The ratings for pain expressed by a female expresser were incorrect for both rater sexes but significantly worse for male raters ($\text{mode}_{\text{pain}} = 0.229$) than for female raters ($\text{mode}_{\text{pain}} = 0.295$). The ratings for pleasure were also incorrect and significantly different between male and female expressers; but, in this case, male raters ($\text{mode}_{\text{pleasure}} = 0.619$) were more correct than the female raters ($\text{mode}_{\text{pleasure}} = 0.534$).
Among the male stimuli, we found that all the affective states were significantly differently rated by male and female participants (Table 2). We found the highest accuracy in rating for low arousal affective states (laugh and neutral) and in both cases female raters ($\text{mode}_{\text{laugh}} = 0.930$; $\text{mode}_{\text{neutral}} = 0.939$) were significantly more accurate than male raters ($\text{lmode}_{\text{laugh}} = 0840$; $\text{mode}_{\text{neutral}} = 0.904$). The ratings for the male expression of pain were incorrectly rated by both the male and female raters in both sexes, but significantly less correct for female raters ($\text{mode}_{\text{pain}} = 0.093$) than for male raters ($\text{mode}_{\text{pain}} = 0.139$). For the expression of pleasure (also rated incorrectly) we found the opposite: female raters ($\text{mode}_{\text{pleasure}} = 0.606$) were more accurate than the males ($\text{mode}_{\text{pleasure}} = 0.536$).

Table 2

Affective States		female stimulus		male stimulus	
		FEMALE raters	MALE raters	FEMALE raters	MALE raters
Pain	mode	0.295	0.229	0.093	0.139
	Confusion matrix	$\begin{pmatrix} 94.9 & 5.1 \\ 6.6 & 93.4 \end{pmatrix}$		$\begin{pmatrix} 93.8 & 6.2 \\ 7.2 & 92.8 \end{pmatrix}$	
	pdtC	0.00	0.00	0.00	0.00
Pleasure	mode	0.534	0.619	0.606	0.536
	Confusion matrix	$\begin{pmatrix} 96.7 & 3.3 \\ 4.1 & 95.9 \end{pmatrix}$		$\begin{pmatrix} 93.8 & 6.2 \\ 7.7 & 92.3 \end{pmatrix}$	
	pdtC	0.058	2.01×10^{-6}	5.95×10^{-7}	0.082
Laugh	mode	0.819	0.832	0.930	0.840
	Confusion matrix	$\begin{pmatrix} 72.9 & 27.1 \\ 44.3 & 55.7 \end{pmatrix}$		$\begin{pmatrix} 99.9 & 0.1 \\ 0.1 & 99.9 \end{pmatrix}$	
	pdtC	3.39×10^{-52}	2.86×10^{-41}	4.64×10^{-102}	2.15×10^{-43}
Neutral	mode	0.855	0.832	0.939	0.904
	Confusion matrix	$\begin{pmatrix} 78.7 & 21.3 \\ 28.1 & 71.9 \end{pmatrix}$		$\begin{pmatrix} 92.2 & 7.8 \\ 8.9 & 91.1 \end{pmatrix}$	
	pdtC	6.07×10^{-66}	2.86×10^{-41}	8.33×10^{-108}	1.76×10^{-63}

Table 2 The modes, the confusion matrices, and the probability due-to-chance (pdtC) of the ratings of the affective states. For probability of due to chance rating we consider the 5% level adequate.

Ratings Due to Chance

One of the advantages of the Bayesian approach is the possibility to test whether the result obtained is consistent ('real' in common parlance) or if it is obtained due to chance. The probability of the result being due to chance (Table 2) ranges between 0 and 0.5; the closer to 0.5, the more probable that the result is due to chance.

Apart from pleasure, the ratings of all of the affective states were rated with a chance probability below 1×10^{-5} (Table 2). We could conclude the ratings for all stimuli except pleasure, were not guessed and the result is reproducible. We highlight again that this does not imply that they were accurate, only that the ratings are not due to guessing and the raters were convinced of their ratings. In the case of pleasure, we found that the ratings of female raters rating the female stimuli ($\mathrm{pdtC_{pleasure}} = 0.058$) and the male participants for male stimuli ($\mathrm{pdtC_{pleasure}} = 0.082$) were not due to chance. This result affects the interpretation of the statistical difference of rating distribution for pleasure: if the stimuli were not rated due to chance, then the differences (if significant) are important.

Discussion

There are numerous ways to produce stimuli for testing. Most often, trained actors and/or actresses are asked to produce facial expressions and vocalizations that are later rated by professionals or naïve respondents in a pre-test. Whenever the within-rater agreement is sufficiently high, the stimulus is used for testing [37].

In our case, this approach is not possible for two reasons: (a) Because our hypothesis postulates that the two expressions that are of highest interest to us (pain and pleasure) are putatively indistinguishable, asking pre-test raters to distinguish these would not be sensible. (b) Using stimuli labeled during the pre-test as pleasure or pain, for example, would inherently lead to testing whether the participating raters agree with the pre-test evaluators on representations of pain and pleasure. That is to say, whether there is a common mental representation as discussed by Chen [38]. The ethological validity of such a result would be extremely limited and has recently been criticized [12, 17, 18].

Instead, we followed the methodology of two pioneering studies on this issue [7, 9]. To do so, we searched for videos online with distinctive facial expressions and vocalizations by expresser(s) — restricting our search to webpages that allowed a free download option. As an extension to the previously mentioned articles, we went one step further in our stimuli choice and insisted on finding (and using) individuals of both biological sexes who each expressed all the five desired facial expressions (together with the simultaneous vocalizations).

Our findings concerning (bimodal) stimuli are in agreement with recent research about affective states: for both sexes, the accuracy probability for low-intensity affective states (laugh/smile and neutral/neutral) is higher than for high-intensity affective states (pain and pleasure).

Male raters showed a tendency to rate more positively the high arousal stimuli produced by female expressers than did the female raters. This tendency has been found to be significant for both pain and pleasure. These results could be explained by a theory of sexual overperception bias, for which male would be more susceptible to false-positive errors when inferring female sexual expressiveness. As an extension of this bias, males may rate female expressions as more positive than female raters. Men generally exhibit a tendency to over-perceive sexual signals [39, 40, 41].

Indeed, when analyzing stimuli produced by male expressers, female raters rated pain and pleasure expressions significantly more positive than did male raters. Importantly, we note that male ratings of male pleasure stimuli and female ratings of female pleasure stimuli are due to chance. The raters were guessing the valence of the stimuli and they were not convinced of their judgment. The due-to-chance analysis is important because, in this case it indicates that even if the ratings are significantly different among the sexes of the raters, they are not reliable: the same participant could rate differently in a hypothetically different moment.

We found no significant differences between male and female raters rating low arousal stimuli produced by female expressers. The raters rated these stimuli with very high accuracy (and not due to chance). Our finding is in agreement with previous studies using the same methodology [12, 17].

Interestingly, for male expressers, we found that the low intensity affective states were significantly better rated by female raters and were not due to chance. Our finding supports the previously suggested female superiority regarding emotion detection and interpretation (females are faster and more accurate in the attribution of the displays; [29, 30, 31, 32, 42]). However, our results do not support a female superiority in expressing the affect by facial expression in isolation (suggested by [41, 42]) nor vocal display in isolation (suggested by [44, 45]). The stimuli of low arousal affective states produced by male expressers showed higher mode component than these same stimuli expressed by females. However, these stimuli (low arousal affective states expressed by males) are rated significantly more accurately when rated by female participants (Table 2).

When we compare the obtained results with the findings in previous studies that presented the same stimuli in an unimodal regime — solely the visual (facial expression) and solely auditory (vocalization) modality — we should be aware that these stimuli were found to be different when using Artificial Intelligence methods [28, 46]; we find the present findings clearly confirm these differences. The main finding of the study on facial expressions in isolation [12] is that the mechanism to distinguish between extremely intensive positive and extremely negative displays (pain and pleasure) has a consistently very low accuracy (among repeated rating tasks), whereas the low-intensity displays (neutral and laugh/smile) are consistently identified with high accuracy. Similar results were obtained for the vocalizations presented in isolation [17], confirming outcomes we obtained in the current study. What differs for high-intensity affective states in the different modalities of presentation is the participants' convictions regarding their choice: while in the case of facial expression the ratings were always due to chance [12] and in the case of vocalization they were never due to chance (even if the decision was not correct, the raters still were convinced of their decision; [17]). Herein, in the case of bimodal stimuli presentation for the expressions of pain, the ratings were never due to chance while for pleasure they were due to chance in only two conditions (female stimuli rated by female raters and male stimuli rated by male raters). These results highlight that, even if the bimodal presentation did not improve the accuracy of the ratings, it changed the level of confidence that the raters had in their answers. Generally, they are more confident than when rating only the facial expression but less confident than when rating only the vocalizations. Furthermore, the findings of our current study supply a strong evidence that the sex of the rater and that of the expresser is important when guessing is involved.

Limitations and Future Directions

One *seeming* limitation would be the prediction of null results. In a statistical sense, it is not to be considered problematic to test for a null effect (but perhaps it was when Null Hypotheses Statistical Testing conventions/paradigms, and the associated fallacies, were widespread). Bayesian statistics is not susceptible to such an issue (because the methods we use do not violate Bayes' Theorem) and specifically includes testing for a null result. Therefore, this approach is promising for future research.

We tested for two types of null results. One null result (often observed): the outcome of a statistical test shows that the observed effect is due to chance. The other type we tested for: that the observed difference of a result that is not due-to-chance but the detected difference is valid with a very small probability.

In both studies presented here, the samples of both stimuli expressers and raters consisted of members of a Caucasian population, since the diversity of the population in the Czech Republic is minimal. The results, although very strong, may not be directly generalizable to other populations.

Female sexual pleasure is objectively difficult to assess, but this difficulty applies to all related research. There are claims that even self-reports would not be sufficient. Devices used for measuring female sexual arousal are insufficiently reliable [47, 48, 49], so we did not use them in this investigation. As in other studies that attempt to relate arousal with female pleasure expression, we use the pragmatic approach: for stimulus creation, it is sufficient to adopt the convention of relying on using already existing, freely downloadable videos with distinctive human vocalizations. Researchers who question this pragmatic approach must then reject the validity of a vast number of studies dealing with vocal expressions of pleasure, not only those using videos. However, it should be pointed out that applying the AI methods to human vocalizations (as was done for facial expressions, [28]) has the potential of resolving this impasse. By the same token, we feel the need to address the possibility that the expression of the stimulus need not match the inner feeling of the expresser. This is not a design flaw but involves an inherently biological aspect in the field of research using naturalistic stimuli.

Lastly, it should be pointed out that the situation of sexual play is not transferable to other types of interaction where such mismatches can be found, e.g., sport, fighting, and injury infliction. Consequently, the generalization of our findings to such fields may have to be used with caution.

Conclusion

The manuscript presents the results of bimodal stimuli presentation of affective states expressions and also compares the outcomes to previous articles that used the same methodology but had only focused on unimodal stimuli presentation. Our study found that the probability of highly accurate assessment is confirmed for low-intensity affective states (laugh/smile and neutral) whereas for high-intensity affective states (pain and pleasure) this is not the case. Furthermore, the male raters tended to rate the bimodal stimuli by female expressers far more positively than female raters, which could be due to a theorized sexual overperception bias found in some previous studies.

We did not identify any significant differences between the male and the female raters rating low arousal stimuli produced by female expressers. However, for male expressers, low-intensity affective states were better rated by female raters. We found no evidence of female superiority in expressing emotions through facial expressions or vocalizations in previous studies, but we did find that, when the bimodal stimuli are used, the low arousal affective states expressed by a male were rated more accurately by female raters.

As a consequence of our findings, we suggest that future studies should focus on the pleasure perception being rated due to chance in case of within-sex rating conditions. Studying this phenomenon may bring important insights into research on sexual consent.

References

1. Donato G, Bartlett MS, Hager JC, Ekman P, Sejnowski TJ. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **21**(10):974–989 (1999).
2. Kibrik, A., & Molchanova, N. Channels of multimodal communication: Relative contributions to discourse understanding. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. **35**, No. 35 (2013).
3. Martinez, L., Falvello, V. B., Aviezer, H., & Todorov, A. Contributions of facial expressions and body language to the rapid perception of dynamic emotions. *Cognition and Emotion*, **30**(5), 939–952 (2016).
4. Zhou, W. & Chen, D. Fear-related chemosignals modulate recognition of fear in ambiguous facial expressions. *Psychological Science*, **20**, 177–183 (2009).
5. Wieser, M. J. & Brosch, T. Faces in context: a review and systematization of contextual influences on affective face processing. *Frontiers in psychology* **3**, 471 (2012).
6. Kayyal, M., Widen, S., & Russell, J. A. Context is more powerful than we think: Contextual cues override facial cues even for valence. *Emotion*, **15**(3), 287 (2015).
7. Aviezer, H., Ensenberg, N., & Hassin, R. R. The inherently contextualized nature of facial emotion perception. *Current Opinion in Psychology*, **17**, 47–54 (2017).
8. Roberts, S. C., Havlíček, J., & Schaal, B. Human olfactory communication: current challenges and future prospects. *Philosophical Transactions of the Royal Society B*, **375**, 20190258 (2020).
9. Aviezer, H., Trope, Y., & Todorov, A. Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, **338**, 1225–1229 (2012).
10. Hughes, S. M. & Nicholson, S. E. Sex differences in the assessment of pain versus sexual pleasure facial expressions. *Journal of Social, Evolutionary, and Cultural Psychology* **2**(4), 289 (2008).
11. Wenzler, S., Levine, S., van Dick, R., Oertel-Knöchel, V., & Aviezer, H. Beyond pleasure

and pain: Facial expression ambiguity in adults and children during intense situations. *Emotion*, **16**(6), 807 (2016).
12. Boschetti, S., Prossinger, H., Hladký, T., Machová, K., & Binter, J. “Eye can’t see the difference”: Facial Expressions of Pain, Pleasure, and Fear Are Consistently Rated Due to Chance. *Human Ethology*, **37**, 046-072 (2022)
13. Atias, D. *et al.* Loud and unclear: Intense real-life vocalizations during affective situations are perceptually ambiguous and contextually malleable. Journal of Experimental *J Exp Psychol Gen.*, **148**(10), 1842 (2019).
14. Holz, N., Larrouy-Maestri, P., & Poeppel, D. The paradoxical role of emotional intensity in the perception of vocal affect. *Scientific reports*, **11**(1), 1-10 (2021).
15. Russell, J. A. A circumplex model of affect. *Journal of Personality and Social Psychology*, **39**(6), 1161 (1980).
16. Posner, J., Russell, J. A., & Peterson, B. S. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, **17**(3), 715–734 (2005).
17. Binter, J., Boschetti, S., Hladký, T., & Prossinger, H. "Ouch!" or "Aah!": Are Vocalizations of 'Laugh', 'Neutral', 'Fear', 'Pain' or 'Pleasure' Reliably Rated?. Preprint at https://psyarxiv.com/rf7vw/ (2023).
18. Van Der Zant, T., & Nelson, N. L. Motion increases recognition of naturalistic postures but not facial expressions. *Journal of Nonverbal Behavior*, **45**(4), 587–600 (2021).
19. Pourtois G, de Gelder B, Vroomen J, Rossion B, Crommelinck M. The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, **11** (6): 1329–1333 (2000).
20. Shimojo S., Shams L. Sensory modalities are not separate modalities: plasticity and interactions. *Curr Opin Neurobiol* **11**(4) 505–509 (2001).
21. de Gelder B, Pourtois G, Weiskrantz L. Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proc Natl Acad Sci USA*. **99** (6): 4121–4126 (2002).
22. Bertelson, P, Aschersleben, G. Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *Int J Psychophysiol*. **50** (1–2): 147–155 (2003).
24. Wada Y., Kitagawa N., Noguchi K. Audio-visual integration in temporal perception. *Int J Psychophysiol.* **50**(1–2) 117–124 (2003).
25. Pourtois G, de Gelder B, Bol A, Crommelinck M. Perception of facial expressions and voices and of their combination in the human brain. *Cortex* **41**(1): 49–59 (2005).
26. Vroomen J., Driver J., de Gelder B. Is cross-modal integration of emotional expressions independent of attentional resources? *Cogn Affect Behav Neuroscience* **1**(4) 382–387 (2001).
27. Belin, P., Fillion-Bilodeau, S., & Gosselin, F. The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior research methods*, **40**(2), 531-539 (2008).
28. Prossinger, H., Hladky, T., Binter, J., Boschetti, S., & Riha, D. Visual Analysis of Emotions Using AI Image-Processing Software: Possible Male/Female Differences between the Emotion Pairs “Neutral”–“Fear” and “Pleasure”–“Pain”. *The 14th PErvasive Technologies Related to Assistive Environments Conference* 342–346 (2021).
29. Rotter, N. G., & Rotter, G. S. Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior*, **12**(2), 139–148 (1988).
30. Thayer, J., & Johnsen, B. H. Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. *Scandinavian Journal of Psychology*, **41**(3), 243–246

(2000).
31. Hall, J. A., & Matsumoto, D. Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, **4**(2), 201 (2004).
32. Hampson, E., van Anders, S. M., & Mullin, L. I. A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behavior*, **27**(6), 401–416 (2006).
33. Dubray, S., Gérard, M., Beaulieu-Prévost, D., & Courtois, F. Validation of a self-report questionnaire assessing the bodily and physiological sensations of orgasm. *The Journal of Sexual Medicine*, **14**(2), 255-263 (2017).
34. Roberson, D., Damjanovic, L., & Kikutani, M. Show and tell: The role of language in categorizing facial expression of emotion. *Emotion Review*, **2**(3), 255–260 (2010).
35. Kruschke, J.K. *Doing Bayesian Data Analysis. A Tutorial with R, JAGS, and STAN*. (2nd Edition. Elsevier/Wiley, Waltham, MA, USA, 2015).
36. Caelen, O. A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, **81**(3), 429–450 (2017).
37. Kätsyri, J., & Sams, M. The effect of dynamics on identifying basic emotions from synthetic and natural faces. *International Journal of Human-Computer Studies*, **66**(4), 233–242 (2008).
38. Chen, C. *et al.* Distinct facial expressions represent pain and pleasure across cultures. *Proceedings of the National Academy of Sciences*, **115**(43), E10013–E10021 (2018).
39. Perilloux, C., Easton, J.A., Buss, D.M. The misperception of sexual interest. *Psychological Science*, **23**(2), 146–151 (2012).
40. Brandner, J. L., Pohlman, J., & Brase, G. L. On hits and being hit on: Error management theory, signal detection theory, and the male sexual overperception bias. *Evolution and Human Behavior*, **42**(4), 331–342 (2021).
41. Younis, I., & Ibrahim, M.A. A beautiful noise: women's coital vocalizations. *Human Andrology* **5**(1), 5–12 (2015).
42. Gregorić, B. *et al.* Recognition of facial expressions in men and women. *Medicina Fluminensis* **50**(4), 454–461 (2014).
43. Hess, U., Blairy, S., Kleck, R.E. The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior*, **21**, 241–257 (1997).
44. Banse, R., & Scherer, K. R. Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, **70**(3), 614 (1996).
45. Scherer, K. R., Banse, R., & Wallbott, H. G. Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural psychology*, **32**(1), 76-92 (2001).
46. Prossinger, H., Hladký, T., Boschetti, S., Říha, D., & Binter, J. Determination of "Neutral"–"Pain","Neutral"–"Pleasure", and "Pleasure"–"Pain" Affective State Distances by Using AI Image Analysis of Facial Expressions. *Technologies*, **10**(4), 75 (2022).
47. Meston, C. M., Levin, R. J., Sipski, M. L., Hull, E. M., & Heiman, J. R. Women's orgasm. *Annual Review of Sex Research*, **15**(1), 173–257(2004).
48. Cooper, E. B., Fenigstein, A., & Fauber, R. L. The faking orgasm scale for women: Psychometric properties. *Archives of Sexual Behavior*, **43**(3), 423-435 (2014).
49. Meston, C. M., & Stanton, A. M. Understanding sexual arousal and subjective-genital arousal desynchrony in women. *Nature Reviews Urology*, **16**(2), 107–120 (2019).

Author Contributions
JB, SB and TH conceptualized the project. JB, SB and TH participated in stimuli creation and the data collection. HP analyzed the data. JB, SB, TH and HP participated in the drafting and

in the revision of the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgments

The study was preregistered on the OSF portal: https://osf.io/bhk6m/.

Funding

This research was funded by the Czech Science Foundation in the project with number GACR 19-12885Y and titled "Behavioral and Psycho-Physiological Response on Ambivalent Visual and Auditory Stimuli Presentation".

Ethics

Institutional Review Board Statement

This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the Faculty of Science, Charles University, Prague, Czech Republic (Protocol Code 2018/08, approval date: 2 April 2018).

Ethics of using isolated audiovisual materials as stimuli

Although the affliction of pain in the context of extreme, yet consensual, sexual activities is typically due to the actions of some other person, the individual experiencing it knows that he/she consented beforehand and has the ability to demand termination at any time while the scenario evolves; he/she thus retains a high degree of control. Typically, a 'safe-word' is used to terminate such physical and/or psychological activities, and, subsequently, after termination, nurturing behaviors are then provided to such an individual. Pain itself is not an aim of the behavior but rather the goal of increasing sensitivity and priming for greater pleasure (in a sexual sense). It is rare that injuries (apart from bruises) are inflicted on the pain-receiving individual. As stated on the production webpage, all participants in the video clips were informed (not by us, but by the directors) about the to-be-filmed scene contents; they agreed to participate and were interviewed by members of the production team after the scene had been completed.

We consider the use of such stimuli as beneficial for science (granted: these stimuli are perceived as controversial by some). They offer the possibility of novel understandings about the problems of the perceptions of human facial displays (accompanied by vocalizations) in several of the evolutionarily most relevant contexts. The acquired knowledge (some of which we have obtained and are presenting in this paper) can, and will certainly be, used in the fields of education, sexuality-related prevention, law enforcement, and therapy. We, therefore, maintain that the benefits by far outweigh the objections to using such stimuli.

Informed consent

An online information text and consent form was supplied. After reading it, a box was to be ticked by each participant (indicating their informed consent) prior to their participation.

Post-study Support and Debriefing

All parts of the design and debriefing were conducted in co-operation with a trained psychologist who also supervised all data collection.

We supplied the participants (after they finished the response process) with a list of contacts: (1) to the principal investigator, (2) to a psychological counseling center, and (3) to an organization that deals with sexuality-related issues.

Data Availability Statement

The data are available on the OSF portal at https://osf.io/yhfv8

The frames were extracted from commercially available online videos. As these videos are proprietary, we can only make the extracted frames we used available from the corresponding author (upon reasonable requests originating from a serious institutional email address).

Conflicts of Interest

The authors declare no conflict of interest.

Appendix I

Bayesian estimation of guessing

Each stimulus is rated as exhibiting one facial and one vocal expression. We do not expect, but do postulate — as a test — that both the facial expression smile (for example) and the vocal expression will be rated positive, while any other rating is considered incorrect. We use a Bayesian approach to determine the maximum likelihood of a correct probability(!). For each stimulus of each facial expression rated by the females (say), let n_1 be the number of ratings that agree with the postulated rating, while n_2 is the number of ratings that disagree with the postulated rating (then $n_1 + n_2 = n$; $n = 526$ for female raters; $n = 376$ for male raters). In Bayesian statistics, in which the probability s is a random variable, the likelihood function, for this situation, thelikelihood function $\mathcal{L}(\mathrm{s}) = pdf(s)$ is the pdf of a Beta Distribution

$$\mathcal{L}(\mathcal{Be}(\alpha,\beta),s) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1}(1-s)^{\beta-1} = \frac{\Gamma(n_1+n_2+2)}{\Gamma(n_1+1)\Gamma(n_2+1)} s^{n_1}(1-s)^{n_2}$$

The probability (in Bayesian statistics) of observing a result disagreeing with the postulate of guessing is then, either

$$\int_0^{\frac{1}{2}} \mathcal{L}(\mathcal{Be}(\alpha,\beta),s)\,ds$$

or

$$\int_{\frac{1}{2}}^{0} \mathcal{L}(\mathcal{Be}(\alpha,\beta),s)\,ds$$

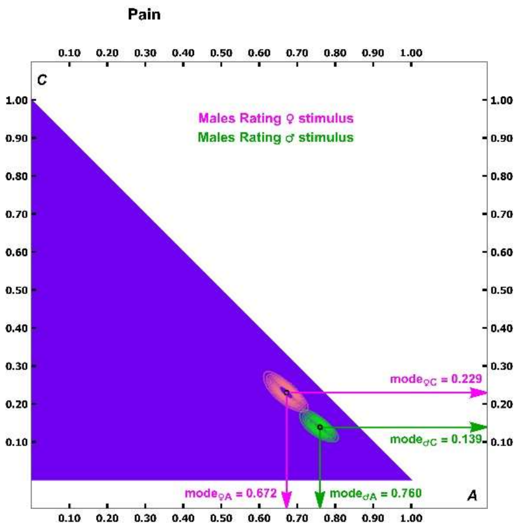
depending on which side of $\frac{1}{2}$ the mode occurs.

The most likely probability s_{ML} is the mode, and for a Beta distribution, $s_{\mathrm{ML}} = \frac{\alpha-1}{\alpha+\beta-2} = \frac{n_1}{n_1+n_2}$. We note that (a) the postulate is always s, even if the postulated rating is negative (as in the case of pain), and (b) these results are independent of n, even though n may be small (for small n, the uncertainty interval $\mathrm{HDI}_{95\%}$ becomes large).

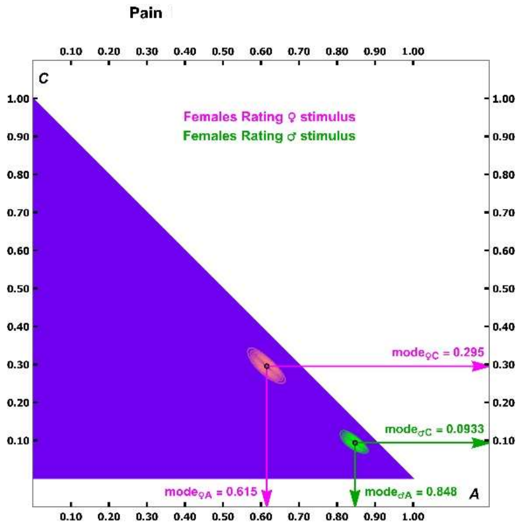
Appendix II

The Dirichlet distributions of **(a)** male raters rating male and female pain stimuli, **(b)** female raters rating male and female pain stimuli, **(c)** male raters rating male and female pleasure stimuli, **(d)** female raters rating male and female pleasure stimuli, **(e)** male raters rating male and female laugh/smile stimuli, **(f)** female raters rating male and female laugh/smile stimuli, **(g)** male raters rating male and female neutral stimuli, **(h)** female raters rating male and female neutral stimuli. The ***A*** axis shows the probability of positive rating; the ***C*** axis shows the probability of negative rating. The domain of each of these Dirichlet functions is shown by the (purple) triangle, because $s_1 + s_2 + s_3 = 1$. In all figures, contour lines are color-coded according to sex of stimulus. Contour lines are in steps of $\frac{1}{14}\mathcal{L}_{\max}$ of the distribution. For neutral, for example, the mode is expected to be very close to $s_2 = 1$; therefore the other two modes should be very close to $s_1 = 0$ and $s_3 = 0$. Fig. 3g and 3h show that this is indeed the case. For laugh, the mode is expected to be very close to $s_1 = 1$; therefore the other two modes should be very close to $s_2 = 0$ and $s_3 = 0$. Fig. 3e and 3f show that this is indeed the case. The corresponding expectations for the modes of pain and pleasure are not met — for neither sex of the raters.

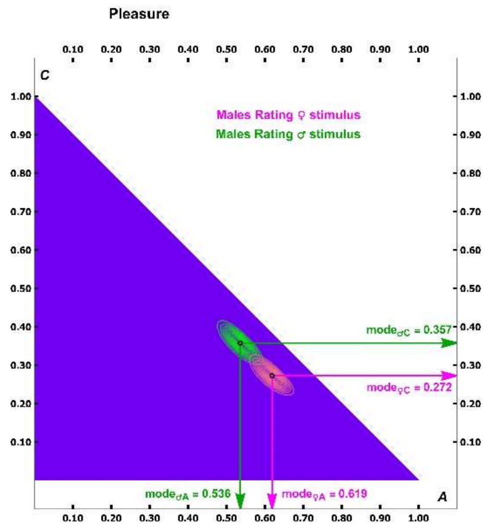
(a)



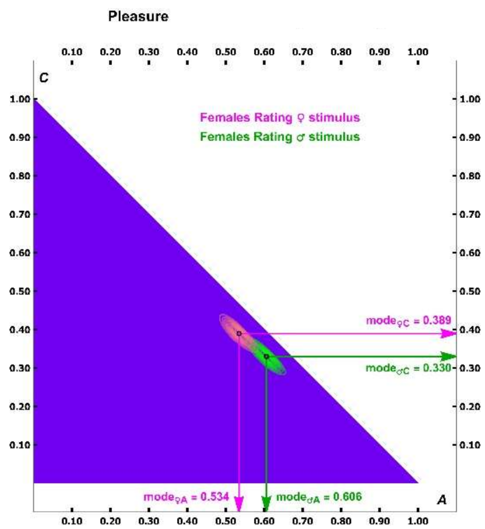
(b)


Pain
C
A
Females Rating ♀ stimulus
Females Rating ♂ stimulus
mode♀C = 0.295
mode♂C = 0.0933
mode♀A = 0.615
mode♂A = 0.848

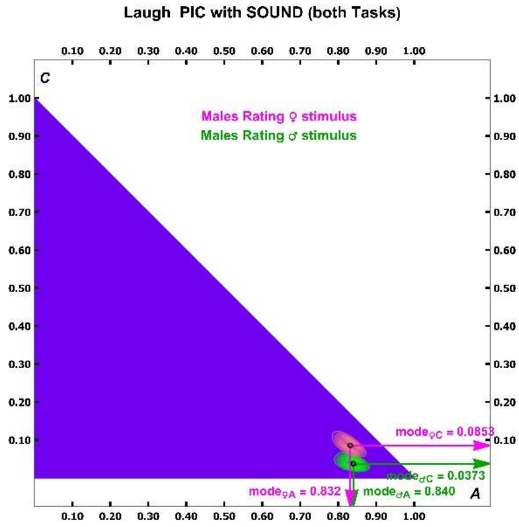
(c)


Pleasure
C
A
Males Rating ♀ stimulus
Males Rating ♂ stimulus
mode♂C = 0.357
mode♀C = 0.272
mode♂A = 0.536
mode♀A = 0.619

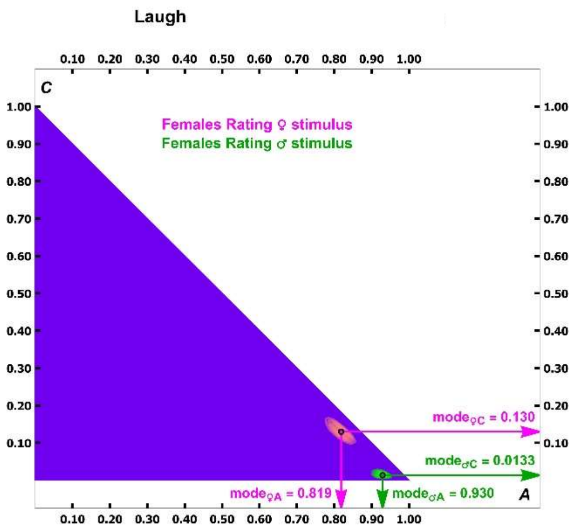
(d)


Pleasure
C
A
Females Rating ♀ stimulus
Females Rating ♂ stimulus
mode♀C = 0.389
mode♂C = 0.330
mode♀A = 0.534
mode♂A = 0.606

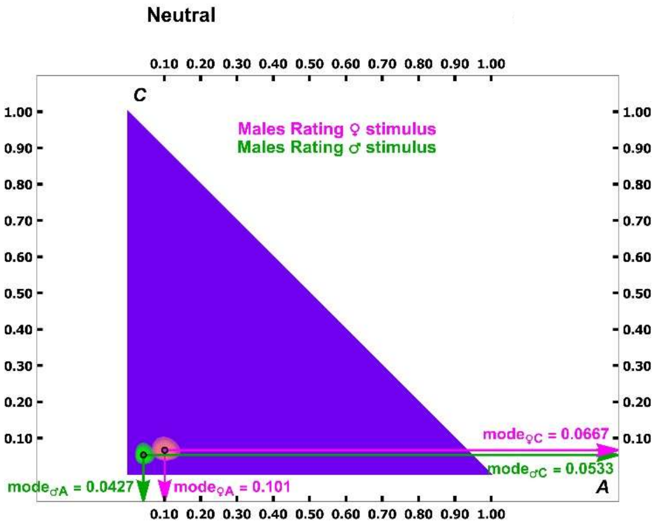
(e)


Laugh PIC with SOUND (both Tasks)
C
A
Males Rating ♀ stimulus
Males Rating ♂ stimulus
mode♀C = 0.0853
mode♂C = 0.0373
mode♀A = 0.832
mode♂A = 0.840

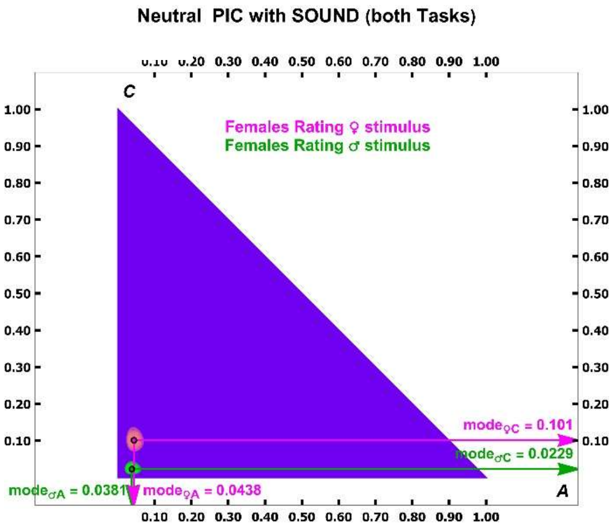
(f)


Laugh
C
A
Females Rating ♀ stimulus
Females Rating ♂ stimulus
mode♀C = 0.130
mode♂C = 0.0133
mode♀A = 0.819
mode♂A = 0.930

(g)


Neutral
C
A
Males Rating ♀ stimulus
Males Rating ♂ stimulus
mode♀C = 0.0667
mode♂C = 0.0533
mode♂A = 0.0427
mode♀A = 0.101

(h)


Neutral PIC with SOUND (both Tasks)
C
A
Females Rating ♀ stimulus
Females Rating ♂ stimulus
mode♀C = 0.101
mode♂C = 0.0229
mode♂A = 0.0381
mode♀A = 0.0438

VOCALIZATION

"Ouch!" or "Aah!": Are Vocalizations of 'Laugh', 'Neutral', 'Fear', 'Pain' or 'Pleasure' Reliably Rated?

Jakub Binter[1], Silvia Boschetti[1,2], Tomáš Hladký[1], & Hermann Prossinger[3]

[1] Faculty of Humanities, Charles University, Prague, Czech Republic

[2] Faculty of Science, Charles University, Prague, Czech Republic

[3] Department of Evolutionary Biology, University of Vienna, Vienna, Austria

*** Corresponding Author**:

Jakub Binter

Faculty of Humanities and Faculty of Science

Charles University,

Prague, Czech Republic

e-mail: jakub.binter@natur.cuni.cz

Key Words: vocalization, emotion, pain and pleasure, Dirichlet distribution, Bayesian statistical approach, Cold Pressor Task

Running head: *Emotional vocalization perception and ratings*

* This manuscript is currently submitted to Human Ethology journal.

Abstract

Our research consisted of two studies focusing on the probability of humans being able to perceive the difference between valence of human vocalizations of high (pain, pleasure and fear) and low intensity (laugh and neutral speech). The first study was conducted online and used a large sample ($n = 902$) of respondents. The second study was conducted in a laboratory setting and involved a stress induction procedure. For both, the task was to categorize whether the human vocalization was rated positive, neutral or negative. Stimuli were audio records extracted from freely downloadable online videos and can be considered semi-naturalistic. Each rating participant (rater) was presented with five audio records (stimuli) of five females and of five males. All raters were presented with the stimuli twice (so as to statistically estimate the consistency of the ratings). Using a Bayesian statistical approach, we could test for consistencies and due-to-chance probabilities. The outcomes support the prediction that the results (ratings) are repeatable (not due to chance) but incorrectly attributed, decreasing the communication value of the expressions of fear, pain,

and pleasure. Stress induction (in study two conducted on 28 participants) did have an impact on the ratings of male neutral and laugh – it caused decrease in correct attribution.

Introduction

Previous studies that deal with acoustic channels

Complex language-based communication is a quite remarkable human characteristic. Thus, paraverbal communication is highly important in social species, such as humans. Much can be deduced from the vocalization e.g., biological sex (Puts et al., 2012), attractiveness (Feinberg et al., 2005), body size (Pisanski & Reby, 2021). Furthermore, prosody plays a vital role in the communication of sexual interest (Hughes & Puts, 2021), of dominance towards listeners (Leongómez et al., 2021), and — importantly — also of affective states (Pisanski et al., 2018).

The communication of affective and emotional states is fundamental for our everyday lives. The interaction among and between humans mainly involves the visual and acoustic channels (Kibrik & Molchanova, 2013). The *concerto* of the information occupying the senses is what creates the final assessment of the communicator's state, being based on context, (linguistic) content, postures, facial expressions and prosody (Leongómez et al., 2022). If the linguistic content is not taken into account, the communicative value of one single signal (i.e. facial expression or vocalization alone) is difficult to extract from the overall multimodal perception.

When only emotional vocalization was rated, the specificity and universality of the vocal production was supported in the case of a negative emotion (Sauter et al., 2010; Gendron et al., 2014) by a cross-cultural study of emotional prosody in both speech (Pell et al., 2009) and emotional vocalization (Gendron et al., 2014). In other studies, the specific role of emotional categories had important implications, with the emotions belonging to the "basic emotions" (anger, fear, disgust, happiness, sadness, and surprise) — presumably because these were more easily recognized in cross-cultural contexts (Bryant & Barrett, 2008; Sauter et al., 2010). Indeed, publications that focused on the communicative value of one single component of the complex expression process showed outcomes that were not in accordance with both major theories of emotions, one conceptualizing emotion as discrete-categorical (Izard, 1994; Ekman & Cordaro, 2011) and one which focused on emotion dimensions (i.e. valence and arousal, Russell, 1980; Posner at al., 2005). Both theories predicted that emotions that are categorically different (discrete category theory) or on opposite sides in terms of valence (emotion dimension theory), are not misunderstood because they are related to very specific

and distinct psychophysiological activation.
Following these theories, we could expect that, even when taken out of context, emotional expressions would be distinguishable among each other's; but data-driven research provides us with counterintuitive results.
Recent results on emotional vocalizations have discovered a novel phenomenon — the misattribution of very intense emotions, called "emotion intensity paradox" (a name given by Holz et al., 2021; even though earlier publications on the topic exist): when the intensity of the emotion is very high, it is more difficult to extract the valence (positive or negative) (Atias et al., 2019).
This misattribution not only occurs for acoustic stimuli; it was previously found in the facial expressions of emotion as well (Aviezer et al., 2012; Hughes & Nicholson, 2008; Wenzler et al., 2016; Boschetti et al., 2022). Facial expressions of emotions of high arousal are not only very difficult to rate correctly, but are oftentimes rated due to chance ('guessing') and inconsistently (Boschetti et al., 2022). The due to chance probability is rarely studied but is a very important metric that allows, based on data distribution, for interpretation related to repeatability of the outcome. It is calculated as the probability by integrating the likelihood function of the Beta distributions over the integral from 0 to $\frac{1}{2}$ or from $\frac{1}{2}$ to 1, depending on which side the mode is; these areas are the probability of the observed distribution of the ratings being due to chance.
A further metric related to the reproducibility is the consistency. The stimuli are presented twice in randomized order; the probability of the same rating being repeated (irrespective of the correctness of the rating) is reported to evaluate how consistent the rated phenomenon is for the raters. It should be pointed out that the two are not dependent on each other, and are not necessarily related to correctness. The ratings may be incorrect, consistent and due to chance or any other combination of the three.
We note that guessing can be inconsistent; therefore, any study dealing with rating issues must test for both guessing and inconsistency. Artificial intelligence (AI) analysis conducted on the facial stimuli identified a further interesting phenomenon. While humans are unable to rate correctly due to inability to extrapolate sufficient cues present in the facial expression, AI can correctly categorize the facial expressions with high accuracy (Binter et al., 2021; Prossinger et al., 2022). The rating inconsistencies by humans is due to their inability to grasp such subtle cues; consequently, the facial expressions of intense emotions are guessed.
This destroys foundations of the communicative value of the extreme affective state as

previously discussed by Aviezer et al. (2012). Since the intensity of the experience (by the expresser) is very high and the situation that evokes it is rich in contextual information regarding the valence, the aim of the behavior may be mainly to capture and orient the attention of those observing (i.e. the receivers). In particular, acoustic stimuli function primarily to gain attention; this has been previously found when comparing screams (intense emotional vocalizations) with regular speech with regard to accuracy and rapidity of localization (Arnal et al., 2015). The results showed that screams were faster and better localized and no differences between natural and synthetic screams (the latter constructed by adding roughness to neutral vocalizations using dedicated software) were found. A recent publication extends this finding and suggests that the ratings of the perceived affects are shifted towards the negative end of the valence scale (Anikin et al., 2020).

Another variable that can have an impact on the success of correct rating is the sex of the rater as well as the sex of the expresser. Previous studies showed that women are better at correct attribution, especially in case of negative emotion evaluation (Thompson & Voyer, 2014). Belin and colleagues (2008) found that male participants rated the expressions as more intense. In that same study, the sex of the expresser was also found to have a significant impact on the ratings of valence and arousal — with a greater arousal but smaller valence attributed to the vocalizations produced by women. The (statistical) association between the sex of the rater and the sex of the expresser was not significant and two groups of authors recommended further studies to clarify this point (Belin et al., 2008; Thompson & Voyer, 2014). In a third study (Vasconcelos et al., 2017), the effect of the sex of the rater was specific for the emotional category (i.e. better recognition of anger and sadness by females in contrast to surprise by males); thus, this (third) study confirms the better recognition of negative expression of vocalizations by women.

There are several limitations in the experimental designs of the majority of the aforementioned studies.

First, because the vocalizations were staged (i.e. often performed by 'neutral' actors who do not elicit the acted emotion, because they did so in a laboratory or in a recording studio). The role of ecological validity clearly emerged as a limiting in case of real versus artificial laugh. The two are not only perceived differently but also cause the activation of different brain regions (McGettigan et al., 2013).

Second, the sex of the raters and the expressers is not always taken into consideration. In light of previous studies (Thompson & Voyer, 2014; Belin et al., 2008; Vasconcelos et al., 2017) showing that this characteristic may systematically affect the correctness of the ratings, it is important to explore the effect of this characteristic — especially the interaction between the sex of the raters and sex of the expressers. Extending on this point, the expressivity of the specific expresser should also be taken into account: some individuals may be more expressive or more stereotypic (or both) in their emotional expression and, therefore, the stimuli produced from such individual may be easier to correctly identify.

Third, a further limitation of the previous studies, which also contributes to the differing and oftentimes contradicting findings, is the presenting of pre-identified emotion categories and/or the use of a rating scale (Bryant, 2021). The pre-identification of categories (i.e. requesting the respondent to choose whether the emotion displayed is anger, fear or surprise) is more likely to capture the complex psychological representation of the emotion and increase the variance, because it could then be more affected by culture and linguistic differences — as criticized by Boschetti at al. (2022). The focus of previous studies was often on the emotions and affects as categories (without attention for the how intense these emotions are) or on the dimensions of the emotions (high vs. low arousal or positive vs. negative), without categorizing the emotions. Consequently, outcomes of a study avoiding these limitations (such as the present one) are very difficult to compare with previous research that focused on the universality of specific categories (such as basic emotions) but not on others (the secondary emotions —the affects).

Fourth, a limitation that should be mentioned is the rater´s state. In the real world scenario, it is almost impossible to be calm and remain in a neutral state while being exposed to a high affect situation that includes the presentation of the extreme vocalizations of pain, pleasure or fear. It was previously shown that the vocalization itself changes (Tolkmitt et al., 1986; Sherer, 2003; Cowie & Cornelius, 2003). Others are also able to differentiate whenever the sender is stressed (Kreiman & Sidtis, 2011; Piskanski et al., 2019) as well as when it can be detected by computer algorithms (Han et al., 2018; Praseito et al., 2019).

There are researches that focused on affect rating while under condition of stress. In all of the previous cases it was facial expression that was rated. One study has identified the shift towards the negative valence of surprised faces (Brown et al., 2017); another study has found this shift in unambiguous faces (neutral and smile) but not in ambiguous faces (pain and pleasure expression) while the physiological changes were found only in response to fear expressions by the same researchers (Boschetti at al., 2022, Binter et al., 2022).

Studies presented in this paper

Our studies are designed to overcome some of the aforementioned methodological limitations. Aim of Study I: to quantify the consistency of the ratings of such vocalizations, focusing on the (biological) sex of the rater as well as the (biological) sexes of the rated (expressers). The stimuli used were audio records of five emotional vocalizations, with high (pain, pleasure and fear) and low (neutral and laugh) intensity. As in previous studies, the stimuli were produced by five different male and five different female expressers (Belin et al., 2008). Each emotion is expressed by five male and five female expressers (so there are 25 stimuli), and each rater rated each of these stimuli twice, presented in random order. We thereby control for the sex and the expressivity of the rated individual (the stimulus expresser).

We generate our set of stimuli by using audio records from videos that depict consensual acts of extreme sexual activities. We adopted a categorical methodology in which there are three ratings: positive, negative, or neutral. We predicted that, if the individual has been exposed to a negative stimulus (pain, for instance) — the vocalization on the audio record (with the expected high intensity), will be rated as negative. In a manifestly opposite stimulus (pleasure, for instance) the rating should be positive. We expect female raters to perform better then male raters in recognizing negative emotions.

Aim of Study II: to evaluate the impact of the physiological state of the rater on his/her ratings. The procedure followed that of Study I with the addition of a stress inducing procedure (Cold Pressure Task; described in more detail below).

Materials & Methods

Sample

Expressers' vocalizations: In order to be consistent with the published terminology, we use the terms "expressers" and "vocalizations" to describe what was presented in the 50 audio records as stimuli. We specify the biological sex of the expresser with the terms male and female. The biological sex of the expressers is documented in the audio records, so we (the authors of this paper) could rely on this information.

Raters: In order maintain consistency with the published terminology, we use the terms "expression raters" and "respondents" to describe the individuals who were presented with the stimuli and who provided their ratings. We specify the biological sex of the expresser with the terms male and female. The biological sex we list is the respondent's self-reported one. We deleted all ratings ($n = 4$) of respondents who did not report their biological sex.

In Study I: A total of 902 individuals (aged 18–50; M_{age} = 32 years, SD = 8.9 years) completed the questionnaires; 526 women (M_{age} = 30.9 years, SD = 8.3 years) and 376 men (M_{age} = 33.6 years, SD = 9.5 year). In Study II, 28 individuals (aged 19–30; M_{age} = 22.3 years, SD = 2.3 years) participated; 13 women (M_{age} = 22.7 years, SD = 2.8 years) and 15 men (M_{age} = 21.9 years, SD = 1.8 years).

Study II: The data were collected in a laboratory in Prague, Czech Republic. 15 participants were presented with the same stimuli as in Study I; the lower right legs of target group members (n = 15) were immersed in cold water (2–4 °C) for 1½ minutes, which subsequently increased their stress level (Cold Pressor Task, CPT; Bullinger et al., 1984; Brown et al., 2017). The control group's 13 participants' lower right legs were immersed in water at room temperature.

Criteria for inclusion were: (a) age of respondents between 18 and 50 years, and (b) at least a minimal experience with adult media, since the vocalizations used in this study were extracted from such materials.

The Two Studies

Study I: The data were collected in the Czech Republic in 2021 via the agency Czech National Panel (narodnipanel.cz) and a science-oriented online portal pokusnikralici.cz using the online platform for data collection Qualtrics®. Participants submitted responses either via computer keyboard or touchscreens of mobile devices (smartphones or tablets).

Study II: The data were collected in a laboratory in Prague, Czech Republic. The participants belonging to the experimental group ($n = 15$) were presented with the same stimuli as in Study I after that the lower right legs had been immersed in cold water (2–4 °C) for 1½ minutes, which subsequently increased their stress level (Cold Pressor Task; Bullinger et al., 1984; Brown et al., 2017). The control group's 13 participants' lower right legs were immersed in water at room temperature before being expose to the stimuli.

Stimuli generation

From the numerous audio-visual materials viewed, ten audio records (five with female vocalizations and five with male vocalizations) were chosen. Based on the plot in each of the audio-visual materials, five vocalizations were selected (one of neutrality, one of fear, one of pleasure, one of pain, and one with laugh). Three of the authors (S.B., J.B. and T.H.) are researchers in field of human sexuality with more than 10 years of experience, specifically focusing on extreme sexual behavior and on consumption of erotic materials. All three authors (one female and two male) provided their opinion on all of the chosen stimuli. Based on the contextual information, all agreed on stimuli chosen and what expression is to be expected. Prior to agreement, stimuli choices were debated among all three researchers in dedicated meetings.

We point out that a common misconception is that the individuals taking part in such exchanges derive sexual pleasures from pain and the two (pain and pleasure) happen simultaneously. Although this may be possible, we have found no mention of it in the published scientific literature. Rather, it should be noted that sensitivity is increased by the experience of pain by various parts of the body (in our case mainly slapping the buttocks and thighs) and only after the painful procedure is over is climax achieved. All the audio stimuli we chose were derived from the whole context of the video. Specifically, we could rely on the images/scenes to identify the emotions and affects (which the raters could not, as they only heard the vocalizations). There is no doubt, due to the camera perspective, about the occurrence of the climax in male expressers. In the female expressers, no such explicit method of judgement can be used, but all signals of the occurrence of climax were identified by the researchers (involving breathing, contraction of pelvic musculature, twitching of anal sphincter muscles, facial blushing, vocalization etc.; Dubray et al., 2017), and further supported by expressers' self-reports at the ends of the videos.

In each audio record, male/female vocalizations expressed fear, pain and pleasure during the session, while laugh and neutral vocalization (speech) were recorded during an interview prior to the pain and pleasure experiences. All stimuli (audio records) were adjusted to the same sound level and lasted from 0.5 seconds to 1.5 seconds — depending on the stimulus.

Procedures

In Study I, the set of stimuli was presented twice (Task 1 and Task 2), each time with a different randomization sequence: each stimulus was played for approximately 1.5 seconds at random intervals ranging from 1 to 3 seconds (so as to avoid constant/rhythmic preparedness for the stimulus presentation). Thus, a total of 100 ratings (two for each of the 50 different stimuli) were collected for each rater.

In Study II, each rater was presented with the set of stimuli (25 male and 25 female vocalizations from five male and five female expressers) only once. The reason is that the Cold Pressor Task (CPT) has limited impact on the cortisol release and this allowed us to finish the procedure within 20 minutes after the effect of the CPT ended.

Ratings

Previous literature (Dolan et al., 2001, Bryant, 2021) has noted that it is a challenging task to correctly identify human vocalizations (e.g., to categorize the expression of fear as indeed fear). We therefore asked our participants to rate the observed expression as either positive, neutral, or negative. We thereby avoided the problem of correct labeling and avoided any intricacies associated with a verbal categorization system. The ratings were communicated either via using keyboard keys or a touchpad with dedicated areas (specified by icons); they were subsequently stored in a dedicated data base.

Statistical Analyses

Due to the inherent advantage of Bayesian statistics when dealing with our research questions, we implemented this approach. General descriptions follow, while more detailed descriptions, augmented by a graphical display, are provided in the Appendix.

(a) Confusion matrices: Both female and male ratings are Dirichlet-distributed (in our case: 3-parametric). The (Bayesian) method of determining whether two groups are significantly different (or not) is to calculate the confusion matrix; it is the obligatory method to use when sample sizes are small. One sample (*F*) has a distribution $dist_F$ and another sample (*G*) has a distribution $dist_G$. When there is an overlap of the *pdf*s (probability density functions) of these two distributions, a fraction of *F* is TRUE (and a fraction is FALSE); likewise, for *G*. The confusion matrix has four entries:

$$\begin{pmatrix} \text{TRUE}_F & \text{FALSE}_F \\ \text{FALSE}_G & \text{TRUE}_G \end{pmatrix}$$

If the off-diagonal elements ($\{\text{FALSE}_F, \text{FALSE}_G\}$) are small, there exists a significant difference between the distributions of *F* and of *G* (the significance level being chosen by the researcher). Observe that the sum of each row in the confusion matrix is $1 = 100\%$. The fractions in the confusion matrix can also be calculated using Monte Carlo methods.

(b) Possibility of effects being due to chance: In Bayesian statistics, the probability s is a random variable ($0 \leq s \leq 1$). The crucial separator for determining chance is $s = \frac{1}{2}$. The probability is either the integral of the likelihood function $\mathcal{L}(s)$ over the interval $0 \leq s \leq \frac{1}{2}$ or the integral over the interval $\frac{1}{2} \leq s \leq 1$, depending on which side of $s = \frac{1}{2}$ the mode is. In either case, the integral determines whether an observation is due to chance. (A graphical description is shown in the Appendix.) We note that the probability due to chance is never greater than 50 %. Since there are positive, neutral, and negative responses, we generate a binary case (the correct responses versus the incorrect responses); then the likelihood function is the probability density function of a Beta distribution (Appendix). For example, for laugh, the correct response is a positive rating while the neutral rating and the negative rating together are incorrect responses.

Results

Probabilities of Correct Ratings

Of the five affective states displayed in vocalizations by each sex, only two were rated with high accuracy (above 85% of correct responses): laugh and neutral (Table 1a and 1b). The laugh vocalizations were correctly assessed by female raters with 0.932 of probability in case of female expressers and with probability of 0.966 in case of male expressers. For the male raters, we observed lower probabilities of correct ratings with stimuli produced either by male or female expressers (0.872 in case of male expressers and 0.893 in case of female expressers); however, the difference between male and female raters was significant only in the case of male expressers (Table 1a and Figure 1a).

The ratings of neutral vocalizations had a very high accuracy (above 90%), independent of the sex of the raters or of the sex of the expressers. Female raters had slightly higher accuracy probability (0.950 for male expressers and 0.926 for female expressers), than the male raters (0.938 for female expressers and 0.912 for male expressers); these differences were not significant (Table 1b and Figure 1b).

For the vocalizations of fear, we observed that the probability of correct ratings by both sexes of raters was very low. Indeed, female raters had only 0.138 probability of correctly rating fear for male expressers and 0.028 for female expressers. For male raters, we observed a probability of 0.184 when expressed by male and 0.224 when expressed by female expressers. These rating probabilities were not significantly different between male and female raters (Table 1c and Figure 1c).

The vocalizations of pleasure also had low probabilities of correct ratings: for female raters the probability of correct rating of pleasure vocalization by female expressers was 0.447 while for male expressers it was 0.442. For male raters the probability of correctly rating pleasure vocalization by female expressers was 0.426 and 0.551 by male expressers. The differences between the probabilities of the ratings by male and female raters were not significant (Table 1d and Figure 1d). We note that these four probabilities are close to the boundary $s = \frac{1}{2}$, so it is important to calculate the due-to-chance probability (the indicator of guessing).

In the case of vocalizations of pain, we observe that the Dirichlet distributions of the ratings are not significantly different for male versus female stimuli — for both the male and the female raters. We also note that, for the female raters, the modes indicate that they rated the stimulus pain incorrectly. For the male raters, we observe that there is no mode of the Dirichlet distribution inside the domain, both for male and female stimuli. The non-existence of a mode necessitates an interpretation of the ratings, guided by the mathematical properties of the Dirichlet distribution. Similar to the graph for the stimulus pleasure (Fig. 1c), it happens that, when the modes have coordinates close to $\frac{1}{2}$ for both correct and incorrect ratings, they (the modes) approach the hypotenuse of the domain. In the case of the pain stimulus, the ML Dirichlet distribution 'pushes' the inferred mode beyond the domain diagonal — the mode therefore no longer exists. Contributing to this non-existence of the mode is the fact that the ratings for correct and incorrect are in the vicinity of $\frac{1}{2}$; because $s_A + s_B + s_C = 1$, the probability of s_B would be forced to be close to zero — if the mode exists inside the domain. In terms of interpreting how this situation can occur, we point out that the (male) raters are not guessing (Table 3). Consequently, they are often rating incorrectly, but they are convinced they are not incorrect; or — phrased differently — their conviction of a correct rating fluctuates. In the case of the stimulus pleasure, this fluctuation is just small enough to ensure the mode remains defined and stays within the domain, but very close to the hypotenuse.

The vocalizations of pain also has a low probability of correct rating. The ratings are actually so incorrect that the mode's *pdf* is forced beyond the hypotenuse and the results are (numerically) invalid. Therefore, the outcome is very similar to the one of the pleasure vocalization ratings where the distribution is almost equally distributed between the extreme poles (Table 1e). There is no significant difference between the ratings provided by the male and female raters nor the ratings of male and female vocalizers.

Table1

(a)

Stimulus	Expresser	Task	Raters	Modes			Confusion Matrix (%)
				positive	neutral	negative	
Laugh	Male	1&2	Male	0.872	0.104	0.024	$\begin{pmatrix} 92.2 & 7.8 \\ 6.2 & 93.8 \end{pmatrix}$ *
	Male	1&2	Female	0.932	0.058	0.010	
	Female	1&2	Male	0.893	0.081	0.026	$\begin{pmatrix} 75.5 & 24.5 \\ 49.2 & 50.8 \end{pmatrix}$
	Female	1&2	Female	0.966	0.025	0.009	

(b)

Stimulus	Expresser	Task	Raters	Modes			Confusion Matrix (%)
				positive	neutral	negative	
Neutral	Male	1&2	Male	0.029	0.938	0.033	$\begin{pmatrix} 71.9 & 28.1 \\ 26.5 & 73.5 \end{pmatrix}$
	Male	1&2	Female	0.031	0.950	0.019	
	Female	1&2	Male	0.061	0.912	0.027	$\begin{pmatrix} 66.5 & 33.5 \\ 30.0 & 70.0 \end{pmatrix}$
	Female	1&2	Female	0.048	0.926	0.026	

(c)

Stimulus	Expresser	Task	Raters	Modes			Confusion Matrix (%)
				positive	neutral	negative	
Fear	Male	1&2	Male	0.653	0.163	0.184	$\begin{pmatrix} 58.0 & 42.0 \\ 28.3 & 71.7 \end{pmatrix}$
	Male	1&2	Female	0.690	0.172	0.138	
	Female	1&2	Male	0.696	0.080	0.224	$\begin{pmatrix} 74.4 & 25.6 \\ 32.9 & 67.1 \end{pmatrix}$
	Female	1&2	Female	0.600	0.073	0.028	

(d)

Stimulus	Expresser	Task	Raters	Modes			Confusion Matrix
				positive	neutral	negative	
Pleasure	Male	1&2	Male	0.404	0.045	0.551	$\begin{pmatrix} 68.4 & 31.6 \\ 57.9 & 42.1 \end{pmatrix}$
	Male	1&2	Female	0.442	0.023	0.535	
	Female	1&2	Male	0.536	0.038	0.426	$\begin{pmatrix} 73.0 & 27.0 \\ 46.3 & 53.7 \end{pmatrix}$
	Female	1&2	Female	0.447	0.005	0.548	

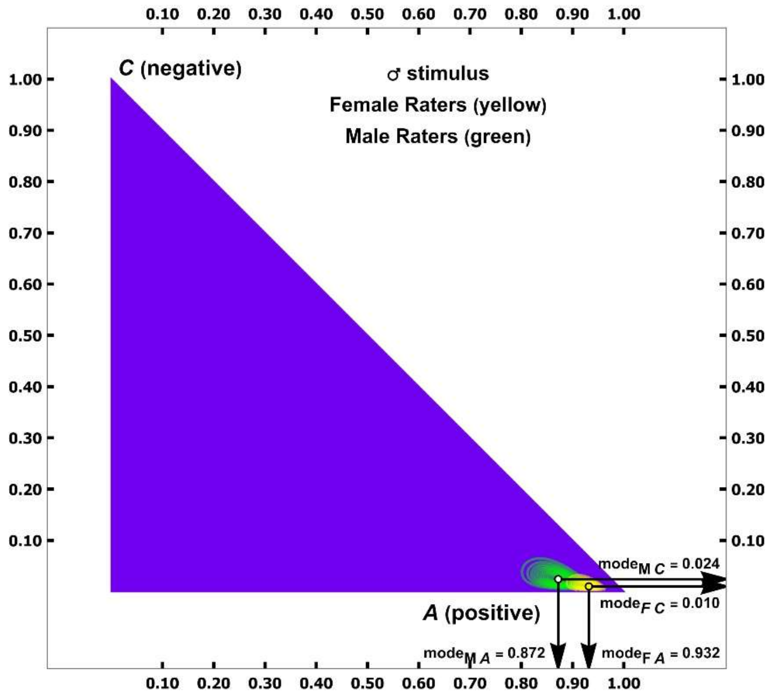
(e)

Stimulus	Expresser	Task	Raters	Modes			Confusion Matrix (%)
				positive	neutral	negative	
Pain	Male	1&2	Male	*nA*	*nA*	*nA*	$\begin{pmatrix} 66.4 & 33.6 \\ 59.2 & 40.8 \end{pmatrix}$
	Male	1&2	Female	*nA*	*nA*	*nA*	
	Female	1&2	Male	0.599	0.099	0.302	$\begin{pmatrix} 80.4 & 19.6 \\ 38.4 & 61.6 \end{pmatrix}$
	Female	1&2	Female	0.521	0.069	0.411	

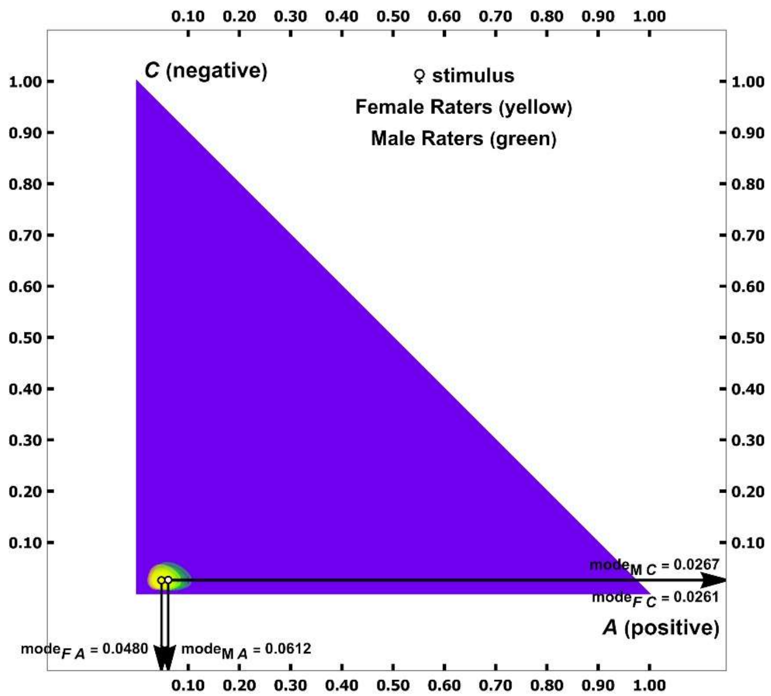
Table 1 The modes and the confusion matrices for the male and female voice stimuli rated by female and male raters. Only for male pain stimulus rated by the female raters (**c**) are the modes outside the domain defined by $s_A + s_B + s_C = 1$ for the Dirichlet distribution, hence expressing a mode is nA (not applicable; further information in the Appendix). If the off-diagonal entries are less than 10% (Caelen, 2017), then the ratings are significantly different; those confusion matrices are marked with an asterisk.

Figure 1

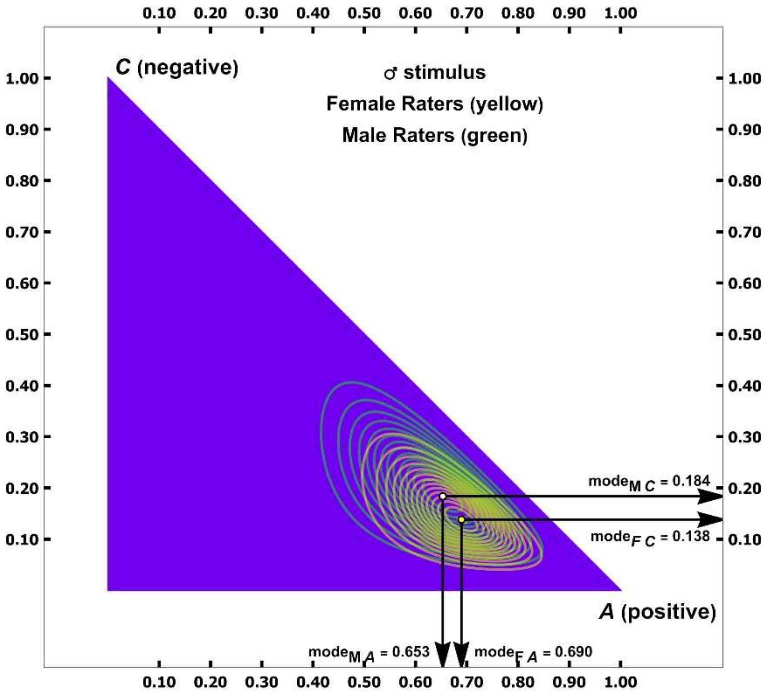
1a Laugh



1b Neutral



1c Fear



1d Pleasure

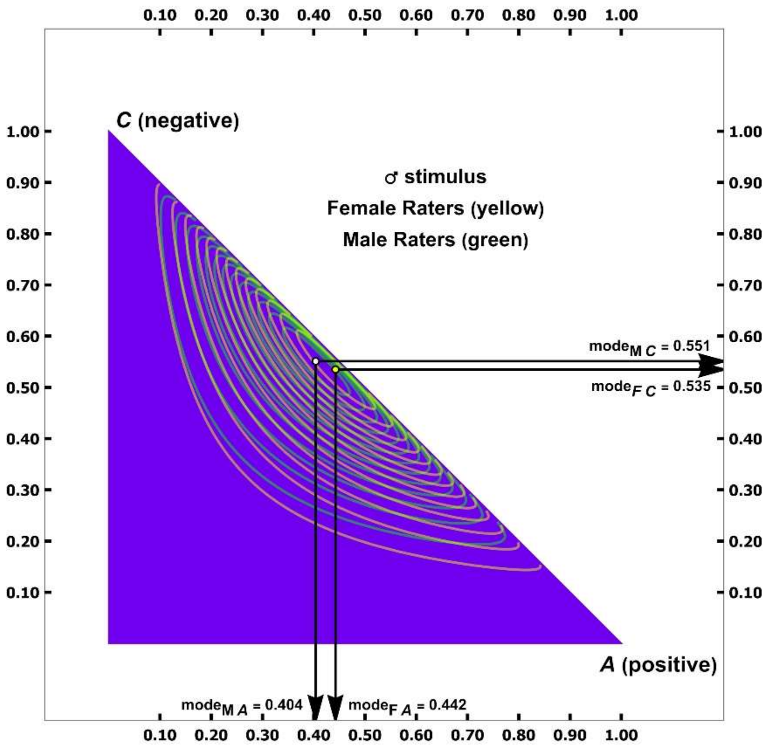


Figure 1 A selection of contour plots of the *pdf*s of the Dirichlet distributions of the ratings of

stimuli by both male and female raters. **(a)** Laugh, male stimulus; **(b)** Neutral, female stimulus; **(c)** Fear, male stimulus; **(d)** Pleasure, male stimulus. In all cases, the *pdf*s of the Dirichlet distributions are defined over the domain (rendered as a purple triangle), because $s_A + s_B + s_C = 1$. Contours are rendered in $\frac{1}{14}$ of the maximum likelihood of the *pdf*. We observe that, the farther the modes are from either $s_A = 1$ or $s_C = 1$, the closer the mode is to the hypotenuse of the domain triangle.

Differences in Ratings by male and female raters

The analyses of male and female differences in correct attributions of an individual expresser (or all expressers) displaying one stimulus (e.g. fear) revealed that all results are not significant. In other words, we confirmed the finding published previously that there is no systematic advantage of one sex correctly rating the presented stimulus over the other. Thus, as is deducible from the further breakdown (Figure 2 and Table 2), there is high degree of similarity between the ratings. Expectedly, neutral and laugh are rated with high assignment accuracy (by both sexes). Interestingly, two male expressers (mA and mB) were also rated with high accuracy by both sexes while the others were rated with equally low probability of correct attribution.

Table 2

Expresser	Confusion Matrix	Affective State	Confusion Matrix
fA	$\begin{pmatrix} 49.0 & 50.9 \\ 43.7 & 56.3 \end{pmatrix}$	Laugh	$\begin{pmatrix} 50.6 & 49.4 \\ 45.2 & 54.7 \end{pmatrix}$
fB	$\begin{pmatrix} 47.4 & 52.5 \\ 40.4 & 59.6 \end{pmatrix}$	Fear	$\begin{pmatrix} 53.2 & 46.8 \\ 43.2 & 56.7 \end{pmatrix}$
fC	$\begin{pmatrix} 50.7 & 49.3 \\ 42.6 & 57.4 \end{pmatrix}$	Pain	$\begin{pmatrix} 51.0 & 48.9 \\ 41.6 & 58.4 \end{pmatrix}$
fD	$\begin{pmatrix} 51.5 & 48.5 \\ 48.6 & 51.4 \end{pmatrix}$	Pleasure	$\begin{pmatrix} 57.3 & 42.7 \\ 49.9 & 50.1 \end{pmatrix}$
fE	$\begin{pmatrix} 53.4 & 46.6 \\ 47.1 & 52.9 \end{pmatrix}$	Neutral	$\begin{pmatrix} 44.7 & 55.3 \\ 44.7 & 55.3 \end{pmatrix}$
mA	$\begin{pmatrix} 49.3 & 50.6 \\ 48.1 & 51.8 \end{pmatrix}$		
mB	$\begin{pmatrix} 50.2 & 49.8 \\ 48.3 & 51.7 \end{pmatrix}$		
mC	$\begin{pmatrix} 51.3 & 48.6 \\ 48.1 & 51.9 \end{pmatrix}$		

mD	$\begin{pmatrix} 51.6 & 48.3 \\ 47.2 & 52.7 \end{pmatrix}$
mE	$\begin{pmatrix} 55.2 & 44.7 \\ 51.0 & 49.0 \end{pmatrix}$

Table 2 The confusion matrices testing whether the male versus female correct ratings of each expresser and of each affective state were significantly different (Fig. 1). The ratings of all five affective states for each expresser are Dirichlet distributions, each with five concentration parameters; the entries in the confusion matrix are $\begin{pmatrix} \text{TRUE}_F & \text{FALSE}_F \\ \text{FALSE}_G & \text{TRUE}_G \end{pmatrix}$.

Figure 2

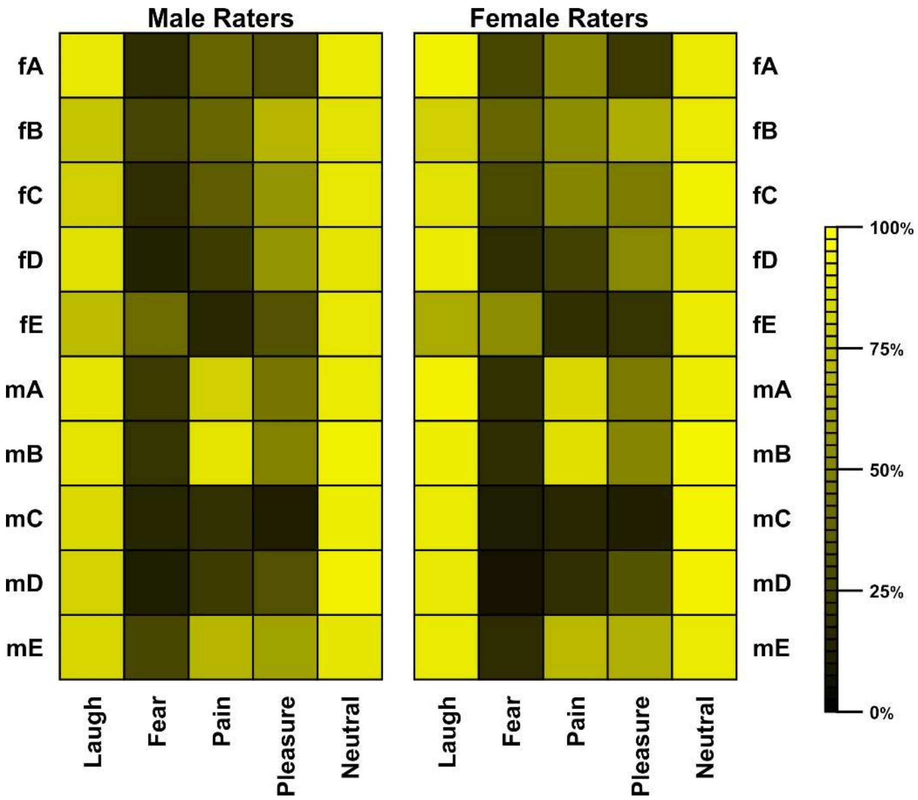


Figure 2 Two heat maps showing the correctness probabilities of ratings by male and female raters of the male vocalizations and the female vocalizations. The male vocalizations and female vocalizations are labeled $\boldsymbol{F}_{\text{index}}$ and $\boldsymbol{M}_{\text{index}}$ respectively.

Ratings Due to Chance

One of the advantages of the Bayesian statistical approach is the possibility to test whether the result obtained is consistent ('real' in common parlance) or if it has been obtained due to chance. The probability of the result being due to chance (Table 3) ranges between 0 and 50%; the closer to 50%, the more probable that the result is due to chance. The ratings of all of the vocalizations were rated with chance probability below one percent. In other words, the rating was not the result of guessing and the result is reproducible. We highlight (again) that this does not mean that the raters are correct or consistent, only that the ratings are not due to guessing and the raters trusted their judgement.

As for patterns that can be deduced and used for further research, it should be pointed out that the female raters were highly consistent when rating neutral vocalizations by men, whereas men were consistent in rating the female fear and male laugh vocalizations (Table 3). Conversely, an extreme inconsistency was found in the case of men rating male fear, women rating both vocalizations of fear and laugh.

Table 3

Stimulus	Raters	DtC_{female}	DtC_{male}
Smile	male	< 1%	< 1%
	female	< 1%	< 1%
Fear	male	< 1%	< 1%
	female	< 1%	< 1%
Pain	male	< 1%	< 1%
	female	< 1%	< 1%
Pleasure	male	< 1%	< 1%
	female	< 1%	< 1%
Neutral	male	< 1%	< 1%
	female	< 1%	< 1%

Table 3 The probabilities that the male and female raters rated the stimuli due to chance (DtC; i.e. the raters were guessing).

Consistency of ratings

Since we presented all stimuli as two consecutively presented tasks, each in a different randomized order, we have the possibility to test the consistency of the ratings. To do so, we have used a Bayesian probability test; the ratings (correct versus incorrect) are Beta distributed.

The confusion matrices that express how significantly different the ratings of the stimuli were between Task 1 (first rating) and Task 2 (second rating) by the female raters (Table 4a) and the male raters (Table 4b). A significant difference is present if both off-diagonal entries are less than 10 % (Caelen, 2017). For example, for the male raters, the first rating of stimulus f_E for laugh was significantly different from the second rating. On the other hand, the first rating (by male raters) of the stimulus m_A for pleasure was not significantly different for the second rating. For the rating of pleasure by the female raters, their ratings were significantly different for $\frac{8}{10}$ of the stimuli. There is no pattern for significant differences of rating of the stimuli, neither by the female nor by the male raters. Because we have evidenced that the ratings are not due to chance (in other words, the raters are not guessing; Table 3), the entries in Table 4 show a remarkable result: even if the raters rate the acoustic stimuli wrongly, they are not guessing; that is to say, they are making a different mistake (wrong rating) when rating again in Task 2. This effect is evident in the *pdf*s of the Dirichlet distributions of Task 1 and Task 2. If the modes are far from the correct rating, then one can be close to the maximal incorrect rating, but also only halfway along the incorrect rating, but then — by definition — close to the neutral rating. In such a scenario, the raters are not guessing, but their incorrect ratings are consistently wrong. Consistently wrong does not mean, however, that they gave the same rating for both tasks. This phenomenon seems to be peculiar to acoustic stimuli. In our publication of visual stimuli, we detected that the raters were guessing (Boscetti et al., 2022), and therefore — by definition — were guessing consistently. In either study (visual or acoustic stimuli): only if the raters were guessing during one task and wrongly rating (but not guessing) during the other task, would the off-diagonal elements be very small. In our case of rating acoustic stimuli, we do not observe this phenomenon. To repeat: we observe that the raters make mistakes (albeit not for every stimulus) — but are not guessing. Very often they made a different rating mistake during Task 1 versus Task 2.

Table 4

(a)

Female Raters					
Expresser	**Laugh**	**Fear**	**Pain**	**Pleasure**	**Neutral**
fA	$\begin{pmatrix} 74.0 & 26.0 \\ 24.1 & 75.9 \end{pmatrix}$	$\begin{pmatrix} 85.8 & 14.2 \\ 15.7 & 84.3 \end{pmatrix}$	$\begin{pmatrix} 90.9 & 9.1 \\ 10.9 & 89.1 \end{pmatrix}$	$\begin{pmatrix} 90.4 & 9.6 \\ 10.5 & 89.5 \end{pmatrix}$	$\begin{pmatrix} 81.0 & 19.0 \\ 21.1 & 78.9 \end{pmatrix}$
fB	$\begin{pmatrix} 91.1 & 8.9 \\ 9.6 & 90.4 \end{pmatrix} *$	$\begin{pmatrix} 91.8 & 8.2 \\ 8.1 & 91.9 \end{pmatrix} *$	$\begin{pmatrix} 99.7 & 0.3 \\ 0.3 & 99.7 \end{pmatrix} *$	$\begin{pmatrix} 99.5 & 0.5 \\ 0.4 & 99.6 \end{pmatrix} *$	$\begin{pmatrix} 52.5 & 47.5 \\ 44.1 & 55.9 \end{pmatrix}$
fC	$\begin{pmatrix} 95.6 & 4.4 \\ 5.6 & 94.4 \end{pmatrix} *$	$\begin{pmatrix} 99.7 & 0.3 \\ 0.2 & 99.8 \end{pmatrix} *$	$\begin{pmatrix} 76.1 & 23.9 \\ 23.8 & 76.2 \end{pmatrix}$	$\begin{pmatrix} 99.7 & 0.3 \\ 0.6 & 99.4 \end{pmatrix} *$	$\begin{pmatrix} 81.9 & 18.1 \\ 16.1 & 83.9 \end{pmatrix}$
fD	$\begin{pmatrix} 92.5 & 7.5 \\ 6.0 & 94.0 \end{pmatrix} *$	$\begin{pmatrix} 97.5 & 2.5 \\ 2.1 & 97.9 \end{pmatrix} *$	$\begin{pmatrix} 99.4 & 0.6 \\ 0.6 & 99.4 \end{pmatrix} *$	$\begin{pmatrix} 99.6 & 0.4 \\ 0.2 & 99.8 \end{pmatrix} *$	$\begin{pmatrix} 94.2 & 5.8 \\ 7.3 & 92.7 \end{pmatrix} *$
fE	$\begin{pmatrix} 99.2 & 0.8 \\ 0.7 & 99.3 \end{pmatrix} *$	$\begin{pmatrix} 96.0 & 4.0 \\ 5.0 & 95.0 \end{pmatrix} *$	$\begin{pmatrix} 97.1 & 2.9 \\ 2.5 & 97.5 \end{pmatrix} *$	$\begin{pmatrix} 88.7 & 11.3 \\ 12.8 & 87.2 \end{pmatrix}$	$\begin{pmatrix} 59.1 & 40.9 \\ 39.2 & 60.8 \end{pmatrix}$
mA	$\begin{pmatrix} 91.1 & 8.9 \\ 7.3 & 92.7 \end{pmatrix} *$	$\begin{pmatrix} 92.2 & 7.8 \\ 7.7 & 92.3 \end{pmatrix} *$	$\begin{pmatrix} 65.2 & 34.8 \\ 35.5 & 64.5 \end{pmatrix}$	$\begin{pmatrix} 95.5 & 4.5 \\ 3.9 & 96.1 \end{pmatrix} *$	$\begin{pmatrix} 90.0 & 10.0 \\ 11.3 & 88.7 \end{pmatrix}$
mB	$\begin{pmatrix} 70.2 & 29.8 \\ 31.9 & 68.1 \end{pmatrix}$	$\begin{pmatrix} 97.1 & 2.9 \\ 3.4 & 96.6 \end{pmatrix} *$	$\begin{pmatrix} 78.9 & 21.1 \\ 22.8 & 77.2 \end{pmatrix}$	$\begin{pmatrix} 99.4 & 0.6 \\ 0.6 & 99.4 \end{pmatrix} *$	$\begin{pmatrix} 88.6 & 11.4 \\ 14.6 & 85.4 \end{pmatrix}$
mC	$\begin{pmatrix} 77.6 & 22.4 \\ 26.3 & 73.7 \end{pmatrix}$	$\begin{pmatrix} 96.4 & 3.6 \\ 3.1 & 96.9 \end{pmatrix} *$	$\begin{pmatrix} 96.0 & 4.0 \\ 3.9 & 96.1 \end{pmatrix} *$	$\begin{pmatrix} 96.6 & 3.4 \\ 4.4 & 95.6 \end{pmatrix} *$	$\begin{pmatrix} 90.3 & 9.7 \\ 10.3 & 89.7 \end{pmatrix}$

mD	$\begin{pmatrix}91.5 & 8.5\\ 11.1 & 88.9\end{pmatrix}*$	$\begin{pmatrix}98.5 & 1.5\\ 1.4 & 98.6\end{pmatrix}*$	$\begin{pmatrix}98.8 & 1.2\\ 1.3 & 98.7\end{pmatrix}*$	$\begin{pmatrix}88.1 & 11.9\\ 12.8 & 87.2\end{pmatrix}$	$\begin{pmatrix}75.0 & 25.0\\ 24.6 & 75.4\end{pmatrix}$
mE	$\begin{pmatrix}93.5 & 6.5\\ 4.8 & 95.2\end{pmatrix}*$	$\begin{pmatrix}98.8 & 1.2\\ 1.1 & 98.9\end{pmatrix}*$	$\begin{pmatrix}91.2 & 8.8\\ 10.7 & 89.3\end{pmatrix}$	$\begin{pmatrix}99.4 & 0.6\\ 0.4 & 99.6\end{pmatrix}*$	$\begin{pmatrix}83.4 & 16.6\\ 15.4 & 84.61\end{pmatrix}$

(b)

Male Raters					
Expresser	**Laugh**	**Fear**	**Pain**	**Pleasure**	**Neutral**
fA	$\begin{pmatrix}89.8 & 10.2\\ 8.3 & 91.7\end{pmatrix}$	$\begin{pmatrix}65.3 & 34.7\\ 36.7 & 63.3\end{pmatrix}$	$\begin{pmatrix}59.7 & 40.3\\ 42.2 & 57.8\end{pmatrix}$	$\begin{pmatrix}80.6 & 19.4\\ 20.9 & 79.1\end{pmatrix}$	$\begin{pmatrix}93.7 & 6.3\\ 8.4 & 91.6\end{pmatrix}*$
fB	$\begin{pmatrix}54.4 & 45.6\\ 43.3 & 56.7\end{pmatrix}$	$\begin{pmatrix}72.9 & 27.1\\ 26.8 & 73.2\end{pmatrix}$	$\begin{pmatrix}91.5 & 8.5\\ 8.7 & 91.3\end{pmatrix}*$	$\begin{pmatrix}81.5 & 18.5\\ 19.2 & 80.8\end{pmatrix}$	$\begin{pmatrix}89.0 & 11.0\\ 11.5 & 88.5\end{pmatrix}$
fC	$\begin{pmatrix}83.0 & 17.0\\ 16.0 & 84.0\end{pmatrix}$	$\begin{pmatrix}89.9 & 10.1\\ 9.4 & 90.6\end{pmatrix}$	$\begin{pmatrix}80.7 & 19.3\\ 20.8 & 79.2\end{pmatrix}$	$\begin{pmatrix}80.3 & 19.7\\ 18.2 & 81.8\end{pmatrix}$	$\begin{pmatrix}96.8 & 3.2\\ 4.5 & 95.5\end{pmatrix}*$
fD	$\begin{pmatrix}89.2 & 10.8\\ 9.6 & 90.4\end{pmatrix}$	$\begin{pmatrix}83. & 17.\\ 16.4 & 83.6\end{pmatrix}$	$\begin{pmatrix}77.4 & 22.6\\ 23.5 & 76.5\end{pmatrix}$	$\begin{pmatrix}93.0 & 7.0\\ 7.4 & 92.6\end{pmatrix}*$	$\begin{pmatrix}76.7 & 23.3\\ 24.8 & 75.2\end{pmatrix}$
fE	$\begin{pmatrix}97.3 & 2.7\\ 2.9 & 97.1\end{pmatrix}*$	$\begin{pmatrix}87.7 & 12.3\\ 13.4 & 86.6\end{pmatrix}$	$\begin{pmatrix}73.5 & 26.5\\ 26.2 & 73.8\end{pmatrix}$	$\begin{pmatrix}87.3 & 12.7\\ 12.6 & 87.4\end{pmatrix}$	$\begin{pmatrix}92.4 & 7.6\\ 9.2 & 90.8\end{pmatrix}*$
mA	$\begin{pmatrix}83.8 & 16.2\\ 16.9 & 83.1\end{pmatrix}$	$\begin{pmatrix}99.9 & 0.1\\ 0.2 & 99.8\end{pmatrix}*$	$\begin{pmatrix}90.9 & 9.1\\ 9.7 & 90.3\end{pmatrix}*$	$\begin{pmatrix}78.3 & 21.7\\ 22.0 & 78.0\end{pmatrix}$	$\begin{pmatrix}94.0 & 6.0\\ 7.4 & 92.6\end{pmatrix}*$
mB	$\begin{pmatrix}90.7 & 9.3\\ 11.7 & 88.3\end{pmatrix}$	$\begin{pmatrix}94.2 & 5.8\\ 5.5 & 94.5\end{pmatrix}*$	$\begin{pmatrix}83.5 & 16.5\\ 15.8 & 84.2\end{pmatrix}$	$\begin{pmatrix}95.6 & 4.4\\ 4.5 & 95.5\end{pmatrix}*$	$\begin{pmatrix}81.2 & 18.8\\ 20.8 & 79.2\end{pmatrix}$
mC	$\begin{pmatrix}65.6 & 34.4\\ 37.8 & 62.2\end{pmatrix}$	$\begin{pmatrix}95.9 & 4.1\\ 3.5 & 96.5\end{pmatrix}*$	$\begin{pmatrix}98.9 & 1.1\\ 0.9 & 99.1\end{pmatrix}*$	$\begin{pmatrix}76.1 & 23.9\\ 23.1 & 76.9\end{pmatrix}$	$\begin{pmatrix}95.1 & 4.9\\ 5.4 & 94.6\end{pmatrix}*$
mD	$\begin{pmatrix}81.1 & 18.9\\ 17.4 & 82.6\end{pmatrix}$	$\begin{pmatrix}94.2 & 5.8\\ 5.5 & 94.5\end{pmatrix}*$	$\begin{pmatrix}98.4 & 1.6\\ 1.6 & 98.4\end{pmatrix}*$	$\begin{pmatrix}89.2 & 10.8\\ 11.7 & 88.3\end{pmatrix}$	$\begin{pmatrix}99.5 & 0.5\\ 0.8 & 99.2\end{pmatrix}*$
mE	$\begin{pmatrix}83.2 & 16.8\\ 16.0 & 84.0\end{pmatrix}$	$\begin{pmatrix}95.7 & 4.3\\ 4.3 & 95.7\end{pmatrix}*$	$\begin{pmatrix}79.5 & 20.5\\ 20.1 & 79.9\end{pmatrix}$	$\begin{pmatrix}83.7 & 16.3\\ 16.9 & 83.1\end{pmatrix}$	$\begin{pmatrix}74.4 & 25.6\\ 25.3 & 74.7\end{pmatrix}$

Table 4 The confusion matrices (entries in %) expressing how consistent the first versus the second ratings of the stimuli are. Significantly different ones are marked with an asterisk. The symbols fA, mark the female stimulus A, mD the male stimulus D, etc.

Stress-induced Rating Differences

In Study II, we analyzed the differences in the distributions of ratings in the two groups of participants (control versus stressed group). The confusion matrices (Table 5) display the probabilities of differences of the ratings by the two groups of participants (separately for male vocalizations and female vocalizations). At a 10% significance level (Caelen, 2017), only two result are significantly different: the laugh and neutral for male vocalizations. The probability of correctness of attribution decreased in the stressed group for laugh from 92.5% of the control group to 81.8%, while for neutral the accuracy decreased from 98.5% to 89.6%. All the remaining ratings are unaffected by the stress induction procedure.

Table 5

Stimulus	Male Vocalizations	Female Vocalizations
Laugh	$\begin{pmatrix} 91.1 & 8.88 \\ 6.82 & 93.2 \end{pmatrix}$ *	$\begin{pmatrix} 53.6 & 46.4 \\ 34.5 & 65.5 \end{pmatrix}$
Fear	$\begin{pmatrix} 67.8 & 32.2 \\ 26.0 & 74.0 \end{pmatrix}$	$\begin{pmatrix} 66.6 & 33.4 \\ 30.1 & 69.9 \end{pmatrix}$
Pain	$\begin{pmatrix} 76.3 & 23.7 \\ 20.9 & 79.1 \end{pmatrix}$	$\begin{pmatrix} 71.7 & 28.3 \\ 24.5 & 75.5 \end{pmatrix}$
Pleasure	$\begin{pmatrix} 68.8 & 31.2 \\ 27.7 & 72.3 \end{pmatrix}$	$\begin{pmatrix} 68.8 & 31.2 \\ 27.7 & 72.3 \end{pmatrix}$
Neutral	$\begin{pmatrix} 95.8 & 4.17 \\ 2.62 & 97.4 \end{pmatrix}$ *	$\begin{pmatrix} 54.6 & 45.4 \\ 33.6 & 66.4 \end{pmatrix}$

Table 5. The confusion matrices between the distributions of the ratings by the control raters and the stressed raters, male vocalizations and female vocalizations separately. Vocalizations that are significantly differently rated are marked with an asterisk.

Discussion

In this paper, we aimed to investigate how isolated intense affective vocalizations are perceived, without context or other associated cues. With our study design, we could overcome some of the limitations of the previously used methodologies. Specifically: (a) We used vocalizations presented in more 'naturalistic' settings (i.e. occurring within an activity as opposed to in a laboratory). (b) For rating responses, we did not request classification of the emotion as a psychological category (i.e. fear, pain) but rather requested the raters to evaluate them using a valence-based rating (as positive, negative, or neutral). (c) It was therefore possible to integrate both theories of emotion (as discrete categories or as dimensional phenomenon) in one experimental setting. (d) The sex of the raters and of the vocalizers were taken into consideration in the study.

The outcomes presented in this paper confirm previous published findings about the emotion intensity paradox: vocalizations of high-intensity affective states (pain, pleasure and fear) are misattributed (and therefore not correctly identified) with very high probabilities. In comparison, the low intensity vocalizations (laugh and neutral) were correctly attributed with high probabilities.

Among the high-intensity affective states, we found that the basic emotion – fear – was more often assigned positive valence (Fig 1c) in contrast to the other two intense affective states that were tested (pain and pleasure). This result confirms — to some degree — the basic emotion theory. While we did not find support for the universality of fear perception (the ratings were incorrect, both for male and female raters — albeit not due to guessing), we did find that the processing of this emotion by the raters was different from the other high-intensity affective states.

An alternative reason why fear may be more misattributed and was predominantly rated as positive (Table 1c) is the specificity of the stimuli used in the current study. The experience of fear may be elicited by an unexpected ‘scary’ or ‘surprising’ situation — in stark contrast to the stimulus we presented. In our case, the expectation of unpleasant experiences that will happen very soon (among them, spanking) seemed not to overly surprise the expressers. In more conventional cases (i.e. the ones most often studied), the emotion of fear is intermixed with surprise — in contrast to the case we studied, in which it is mixed with an anticipatory anxiety. The vocalization mainly consists of intense breathing and soft weeping (connected as it is with anxiety) and therefore the stimulus may be perceived as more positive by the raters, This may be especially misleading when other positive but difficult-to-categorize vocalizations are present. Further studies contrasting vocalizations of both these types of fear (scream as a result of a sudden, fear-inducing emotion — such as a scene in a horror video (Prossinger et al., 2021) versus vocalizations of fear due to anticipatory anxiety —as just before a bungee jump) would contribute to clarifying this issue.

The results for pain vocalizations are also in agreement with previous findings; it was the least correctly rated affective state among all the stimuli presented (Anikin et al., 2017; Lima et al., 2013; Belin et al., 2008).

On the other hand, perhaps unexpectedly, the results for pleasure vocalizations contradict several previously presented study outcomes in which pleasure was either well-recognized (Lima et al., 2013; 2014) or at least correctly attributed to the positive valence rating (Belin et al., 2008). However, our results are in agreement with the studies on vocalization of intense affective states and with previous studies on emotional facial expressions, further confirming that for stimuli with high intensity it is more difficult to extrapolate the correct valence from one single, isolated component (i.e. exclusively vocalization or exclusively facial expression). This interpretation is further supported by the very different results we found for low intensity affective states, showing that our participants correctly assessed the valence of these types of stimuli (Table 1a and 1b).

One argument for the phenomena being counterintuitive is that in such highly intense emotional states, the context (i.e. what elicit the emotion) very clearly points to the valence of the emotion, and, consequently, the affective expression itself need not convey information regarding its valence but rather bring attention to the stimulus within its context. In most of these close-to-natural scenarios, the context would provide sufficient further information that would then contextualize the (vocal and facial) expressions, enabling ratings of as positive or negative.

One natural scenario in which the context may not be of assistance to correctly assess the expression (in terms of facial expression or/and vocalization) is sexual intercourse. Indeed, in the sexual intercourse situation, the stimulus (for example, penetration) may lead to either a positive (pleasure) or a negative (pain) affective state and to the display such an affective state.

One further ramification of the present study is that it highlights how our intuitive interpretation may be in error (as do some others dealing with the emotional intensity paradox; Holz et al., 2021; Atias et al., 2019). As pointed out by Boschetti at al. (2022) for facial cues, this present study brings attention to the possibility of misunderstanding of cues, especially in the above-mentioned situations. The misattribution in these contexts can be avoided when they are accompanied with clarifying verbal communications.

Our novel analytical approach allows us to investigate not only the correct versus incorrect ratings of the stimuli, but also the probability that the ratings could be due to chance: the raters could be guessing, but correctly guessing (or incorrectly guessing). We find that the participant's ratings are not due to chance. When rating the emotional vocalizations, the participants are not guessing the valence (positive, neutral or negative) but they are convinced of the valence of their rating, even when they are incorrectly rating. In another study in which the due-to-chance probability of rating facial expression perception was analyzed (Boschetti et al., 2022), the findings were very different. When rating facial expressions of intense emotions, the participants guessed the valence (specifically — in contrast to the findings presented in this paper) wrongly. When the participants rated the vocalizations (while not being able to see the facial expression), the participants did not guess; they were convinced of the valence of their (wrong) ratings.

A further insight we gained while researching the vocalizations is the consistency of their rating. It is rarely studied even though it should constitute one of the fundamental questions. Is the rating repeatable? We had this expectation for our stimuli; interestingly, only few vocalizations were consistently rated (Table 4). Surprisingly, the female raters were highly consistent in ratings of neutral vocalizations by men but not by women — whereas men were consistent in rating the female fear and male laugh vocalizations. Conversely, an extreme inconsistency was found in case of men rating male fear, women rating fear and laugh. All three of these outcomes draw into question the classical concept of basic emotion perception. (Especially the just-so-stories about the female greater ability to assess the positive affects).

As in the case of facial expressions using the same methodology (Boschetti et al., 2022) the due-to-chance analyses provided a novel tool to study affect perceptions. We found that facial expressions (other than laugh and neutral) are rated due to chance. The results for vocalizations are stunningly different. None of the stimuli, no matter how ambiguous, was rated with uncertainty on the part of the raters. This shows that there is a high reliance on the acoustic perception when compared to the visual perception in case of affect perception.

It was expected that women would be better at correct attribution, especially in case of negative emotion evaluation (Thompson & Voyer, 2014; Belin et al., 2008). In a study by Vasconcelos et al. (2017), the effect of the sex of the rater was specific for the emotional category (i.e. better recognition of anger and sadness vocalization by females in contrast to surprise by males). Our result is in disagreement with both these previous studies; we did not identify any advantage on the side of any sexes in the attribution accuracy, nor on the expresser's sex effect. Nor did we find any support for the finding that some vocalization category was better identified.

We conclude that the intrasexual variation was high on the side of the vocalizers. Some male and some female expressers were rated more accurately than the others (Table 4). This should be further studied by including a possible similar effect on the side of the rater. It can be expected that there are individuals with a higher ability to differentiate the vocalizations. While it exceeds the scope of this article, clustering algorithms are a viable way to identify sub-groups of individuals through multiple assessments (accuracy, consistency, and due-to-chance ratings). The influence of stress on rating of vocalizations is a unique feature of our study. None of the results are due-to-chance and the shift only occurred in case of the non-ambiguous vocalizations (neutral and laugh). The direction is always towards the lower accuracy of ratings and only for the male expressers. We do not have an interpretation for this result; it is the first time it is presented, so we cannot compare with published studies. Thus in similar study focused on the facial expressions, it was pleasure and smile of the male expressers that were rated more accurately by the stressed group . The inner state (stress induction) alters neither the female facial nor vocal ratings. The male vocalizations may be perceived in an altered way as caution for dangers. Again, further studies would be necessary to extend our knowledge on the topic.

Limitations and Future Directions

One *seeming* limitation of any study would be the prediction of null results. In a statistical sense, it is considered problematic to test for a null effect (but that may perhaps be due to Null Hypotheses Statistical Testing conventions/paradigms and the associated fallacies). Bayesian statistics is not susceptible to such a problem (because the Bayesian methods do not violate Bayes' Theorem) and it specifically includes testing for a null result. Therefore, this approach is promising for future research.

We tested for two types of null result. One null result (often observed): the outcome of a statistical test shows that the observed effect is due to chance. The other type we tested for: that the observed difference of a result that is not due to chance but the detected difference is valid with a very small probability.

In both studies presented here, the samples of both stimuli and raters consisted of members of a Caucasian population, since the diversity of population in Czech Republic is minimal. The results, although very strong, may not be directly generalizable to other populations.

Female sexual pleasure is objectively difficult to assess; but this difficulty applies to all related research. There are claims that even self-reports would not be sufficient. Devices used for measuring female sexual arousal are insufficiently reliable (Meston et al., 2004; Cooper et al., 2014; Meston et al., 2019), so we did not use them in this investigation. As in other studies that attempt to relate arousal with female pleasure expression, we use the pragmatic approach: for stimulus creation, it is sufficient to adopt the convention of relying on using already existing, freely downloadable videos with distinctive human vocalizations. Researchers who question this pragmatic approach must then reject the validity of a vast number of studies dealing with vocal expressions of pleasure, not only those using videos. However, it should be pointed out that applying the AI methods to human vocalizations (as was done for facial expressions, Prossinger et al., 2022) have the potential of resolving this impasse.

By the same token, we feel the need to address the possibility that the expression heard does not match the inner feeling of the expresser. This is not a design flaw but involves an inherently biological aspect in the field of research using naturalistic stimuli.

Furthermore, the expression of fear as a reliable stimulus may be considered problematic since the expressers were aware of the fact that, ultimately, the situation is safe: no permanent damage is de facto guaranteed by the plot of the video. Fear, of all expressions considered basic, has the lowest identification reliability rate, and this is especially true in naturalistic expression scenarios. In other words, the results obtained are less unusual than may appear at first glance.

Lastly, it should be pointed out that the situation of sexual play is not transferable to other types of interaction in which such mismatches can be found, e.g., sport, fighting, injury infliction. Therefore, generalization of our findings to include such fields may have to be used with caution.

Conclusion

These two studies we presented here bring multiple novel insights to vocalization perception. The first study, with a large sample of participants (exceeding 900) in combination with novel analytical approaches provide us with numerous findings. The low arousal vocalizations (laugh and neutral state) are rated with very high accuracy whereas the fear, pain and pleasure are not. The ratings of the pain are so scattered (extremely high variability) that it is impossible to assign them a mode whereas the rating of pleasure is almost equally distributed on the extreme poles of the rating distribution making both of these vocalizations rated with insufficient accuracy. Fear was highly mistaken for positive vocalization; it can be interpreted for this specific type of situation where surprise is not involved.

We found no sex differences between the vocalizers or the raters to have an impact — with one exception. There is a pattern of consistency rating where female raters were consistent in ratings of neutral vocalization by men but not by women whereas male raters were consistent in rating the female fear and male laugh. Conversely, an extreme inconsistency was found in the case of men rating male fear, women rating fear and laugh.

None of the ratings were due to chance; actually the probability of the rating being due-to-chance (guessing) was smaller than one percent. The ratings were often incorrect and inconsistent, but they are not the result of guessing.

The second study showed shifts in two male vocalizations – laugh and neutral, after a stress induction procedure. Ratings for both these vocalizations were less accurate in the stressed group.

These many outcomes provide further support for the emotion intensity paradox yet also undermine some of the core concepts of emotional vocalization research. Further studies will be necessary to uncover more about the phenomena we discovered; especially using naturalistic and semi-naturalistic ecologically valid stimuli so as to avoid pre-tested and laboratory obtained stimuli We recommend that even those studies that rely on stimuli dataset (with known previous results) should be tested using the novel statistical methods provided herein. Lastly, the Cold Pressor Task is an ideal stress induction method; there is a lack of studies in which the arousal of respondents is altered; arguably, a key question related to ecological validity.

Ethics

Although the materials presented to the participants were not *per se* of a sexual nature (as only audio records were presented), we made precautions to limit any negative impact on our raters.

Informed consent

In Study I: An online information text and consent form was supplied; after reading it, a box was to be ticked by each participant (indicating their informed consent) prior to their participation.

In Study II: Two informed consent forms were to be manually/personally signed. The first was presented to a potential rater prior to participation; it included all the information about procedures (including the CPT), safety measures, kinds of data collected, and risks. The second informed consent form consisted of a full disclosure of the aim(s) of the study, the expected impact of the procedures, and the possible implications for the rater signing this second form. It was to be signed after the debriefing procedure (see below). If the second consent form was not signed, the collected data was discarded (and therefore not used in the analysis).

Post-study Support and Debriefing

All parts of the design and debriefing were conducted in co-operation with a trained psychologist who also supervised all data collection.

For Study I we supplied the participants with a list of contacts: (1) to the principal investigator, (2) to a psychological counseling center, and (3) to an organization that deals with sexuality-related issues.

During the debriefing phase for Study II, every rater participated in a debriefing discussion by a trained psychologist directly after the completion of data collection. The rater then received a written detailed description, with a full explanation of the possible negative aspects of the experiment, especially those related to the stress-induction procedure, and was also supplied with a list of contacts: (1) to the principal investigator, (2) to a psychological counseling center, and (3) to an organization that deals with sexuality-related issues.

Acknowledgments

The study was preregistered on OSF portal: https://osf.io/bhk6m/.

Funding

This research was funded by the Czech Science Foundation in the project with number GACR 19-12885Y and titled “Behavioral and Psycho-Physiological Response on Ambivalent Visual and Auditory Stimuli Presentation”.

Institutional Review Board Statement

This study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Faculty of Science, Charles University, Prague, Czech Republic (Protocol Code 2018/08, approval date: 2 April 2018).

Data Availability Statement

The data are available on the OSF portal.

The frames were extracted from commercially available online videos. As the videos are proprietary, we can only make the extracted frames we used available from the corresponding author (upon reasonable requests originating from a serious institutional email address).

Conflicts of Interest

The authors declare no conflict of interest.

References

Anikin, Andrey; Persson, Tomas (2017). Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods*, 49(2), 758–771. doi:10.3758/s13428-016-0736-y

Anikin, A., Pisanski, K., & Reby, D. (2020). Do nonlinear vocal phenomena signal negative valence or high emotion intensity? *Royal Society open science*, 7(12), 201306.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology*, 25(15), 2051–22056.

Atias, D., Todorov, A., Liraz, S., Eidinger, A., Dror, I., Maymon, Y., & Aviezer, H. (2019). Loud and unclear: Intense real-life vocalizations during affective situations are perceptually ambiguous and contextually malleable. *Journal of Experimental Psychology: General*, 148(10), 1842.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, 338(6111), 1225–1229. doi: 10.1126/science.1224313

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior research methods*, 40(2), 531–539.

Binter, J., Boschetti, S., Hladký, T., Prossinger, H., Wells, T. J., Jílková, J., & Říha, D. (2022). Quantifying the Rating Performance of Ambiguous and Unambiguous Facial Expression Perceptions Under Conditions of Stress by Using Wearable Sensors. *In International Conference on Human-Computer Interaction (*pp. 519–529). Springer, Cham.

Boschetti, S., Prossinger, H., Hladký, T., Machová, K., & Binter, J. (2022). "Eye can't see the difference": Facial Expressions of Pain, Pleasure, and Fear Are Consistently Rated Due to

Chance. *Human Ethology*, Volume 37, 46–72.

Brown, C. C., Raio, C. M., & Neta, M. (2017). Cortisol responses enhance negative valence perception for ambiguous facial expressions. *Scientific Reports*, 7(1), 1–8. doi: 10.1038/s41598-017-14846-3

Bryant, G., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, 8(1-2), 135-148.

Bryant, G. A. (2021). The evolution of human vocal emotion. *Emotion Review*, 13(1), 25–33.

Bullinger, M., Naber, D., Pickar, D., Cohen, R. M., Kalin, N. H., Pert, A., & Bunney Jr, W. E. (1984). Endocrine effects of the cold pressor test: relationships to subjective pain appraisal and coping. *Psychiatry Research*, 12(3), 227–233. doi: 10.1016/0165-1781(84)90028-3.

Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, 81(3), 429–450.

Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech communication*, 40(1–2), 5–32.

Dolan, R. J., Morris, J. S., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences,* 98(17), 10006–10010.

Dubray, S., Gérard, M., Beaulieu-Prévost, D., & Courtois, F. (2017). Validation of a self-report questionnaire assessing the bodily and physiological sensations of orgasm. *The Journal of Sexual Medicine*, 14(2), 255–263.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic? *Emotion Review*, 3, 364 –370. http://dx.doi.org/10.1177/1754073911410740

Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Smith, M. J. L., Cornwell, R. E., ... & Perrett, D. I. (2005). The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, 26(5), 398–408.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Cultural relativity in perceiving emotion from vocalizations. *Psychological science*, 25(4), 911–920.

Han, H., Byun, K., & Kang, H. G. (2018, October). A deep learning-based stress detection algorithm with speech signal. *In proceedings of the 2018 workshop on audio-visual scene understanding for immersive multimedia* (pp. 11–15).

Holz, N., Larrouy-Maestri, P., & Poeppel, D. (2021). The paradoxical role of emotional intensity in the perception of vocal affect. *Scientific reports*, 11(1), 1–10.

Hughes, S. M. & Nicholson, S. E. (2008). Sex differences in the assessment of pain versus sexual pleasure facial expressions. *Journal of Social, Evolutionary, and Cultural Psychology* 2(4), 289. doi: 10.1037/h0099338

Hughes, S. M., & Puts, D. A. (2021). Vocal modulation in human mating and competition. *Philosophical Transactions of the Royal Society B*, 376(1840), 20200388.

Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research. *Psychological Bulletin*, 115(2), 288–299. doi: 10.1037/0033-2909.115.2.288.

Kibrik, A., & Molchanova, N. (2013). Channels of multimodal communication: Relative contributions to discourse understanding. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 35, No. 35).

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception.* John Wiley & Sons.

Leongómez, J. D., Havlíček, J., & Roberts, S. C. (2022). Musicality in human vocal communication: an evolutionary perspective. *Philosophical Transactions of the Royal Society B*, 377(1841), 20200391.

Leongómez, J. D., Pisanski, K., Reby, D., Sauter, D., Lavan, N., Perlman, M., & Varella Valentova, J. (2021). Voice modulation: from origin and mechanism to social impact. *Philosophical Transactions of the Royal Society B*, 376(1840), 20200386.

Lima, C. F., Alves, T., Scott, S. K., & Castro, S. L. (2014). In the ear of the beholder: how age shapes emotion processing in nonverbal vocalizations. *Emotion,* 14(1), 145.

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior research methods*, 45(4), 1234–1245.

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., & Scott, S. K. (2013). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral Cortex*, 25(1), 246–257.

Pisanski, K., Kobylarek, A., Jakubowska, L., Nowak, J., Walter, A., Błaszczyński, K., ... & Sorokowski, P. (2018). Multimodal stress detection: Testing for covariation in vocal, hormonal and physiological responses to Trier Social Stress Test. *Hormones and Behavior*, 106, 52–61.

Pisanski, K., & Reby, D. (2021). Efficacy in deceptive vocal exaggeration of human body size. *Nature communications*, 12(1), 1–9.

Pisanski, K., & Sorokowski, P. (2021). Human stress detection: cortisol levels in stressed speakers predict voice-based judgments of stress. *Perception*, 50(1), 80–87.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of*

Phonetics, 37(4), 417–435.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(3), 715–734.

Prasetio, B. H., Tamura, H., & Tanno, K. (2019, October). A deep time-delay embedded algorithm for unsupervised stress speech clustering. In: *2019 IEEE international conference on systems, man and cybernetics (SMC)* (pp. 1193–1198).

Prossinger, H., Hladky, T., Binter, J., Boschetti, S., & Riha, D. (2021b). Visual Analysis of Emotions Using AI Image-Processing Software: Possible Male/Female Differences between the Emotion Pairs "Neutral"–"Fear" and "Pleasure"–"Pain". *The 14th PErvasive Technologies Related to Assistive Environments Conference* (pp. 342–346). doi: 10.1145/3453892.3461656

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. P*roceedings of the Royal Society B: Biological Sciences*, 279(1728), 601–609.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161. doi: 10.1037/h0077714

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, 107(6), 2408–2412.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech communication*, 40(1-2), 227–256

Thompson, A. E.; Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion*, 28(7), 1164–1195. doi:10.1080/02699931.2013.875889

Tolkmitt, F. J., & Scherer, K. R. (1986). Effect of experimentally induced stress on vocal parameters. *Journal of Experimental Psychology: Human Perception and Performance*, 12(3), 302.

Vasconcelos, M., Dias, M., Soares, A. P., & Pinheiro, A. P. (2017). What is the melody of that voice? Probing unbiased recognition accuracy with the Montreal affective voices. *Journal of Nonverbal Behavior*, 41(3), 239–267.

Wenzler, S., Levine, S., van Dick, R., Oertel-Knöchel, V., & Aviezer, H. (2016). Beyond pleasure and pain: Facial expression ambiguity in adults and children during intense situations. *Emotion*, 16(6), 807. doi: 10.1037/emo0000185

Appendix

Dirichlet Distribution

The ratings by female raters are Dirichlet distributions (in our case with three concentration parameters $\{\alpha_A, \alpha_B, \alpha_C\}$), as are those of the males. We predicted the repeats (Trial I versus Trial II) to be the same, and we tested for that. We therefore have, for female raters rating five female fear vocalizations, ten registration sets with triples $\{n_A, n_B, n_C\}$ in each set, with $n_A + n_B + n_C = 10$. The *pdf* (probability density function) of the Dirichlet distribution $\mathcal{D}ir$, called the likelihood function $\mathcal{L}(s_1, s_2, s_3) = pdf(\mathcal{D}ir(\alpha_A, \alpha_B, \alpha_C), s_1, s_2, s_3)$, with concentration parameters $\{\alpha_A, \alpha_B, \alpha_C\}$ and probabilities s_1, s_2, s_3 of observing the variables var_1, var_2, var_3 is

$$\mathcal{L}(s_1, s_2, s_3) = pdf(\mathcal{D}ir(\alpha_A, \alpha_B, \alpha_C), s_1, s_2, s_3) = \frac{\Gamma(\alpha_A + \alpha_B + \alpha_C)}{\Gamma(\alpha_A)\Gamma(\alpha_B)\Gamma(\alpha_C)} s_1^{\alpha_A - 1} s_2^{\alpha_B - 1} s_3^{\alpha_C - 1}$$

with $s_3 = 1 - s_1 - s_2$ and $0 \leq s_i \leq 1 \; \forall i \; i = 1 \ldots 3$; $\Gamma(\cdots)$ is the Gamma function.

The two modes for A and C are $mode_A = \frac{\alpha_A - 1}{\alpha_A + \alpha_B + \alpha_C - 3}$ and $mode_C = \frac{\alpha_C - 1}{\alpha_A + \alpha_B + \alpha_C - 3}$. If we are interested in axes A and B, rather than A and C, then the formulae are cycled. In the text, we justify why we use which axes and when. Note that the formulae for the modes are straightforward, suggesting we need not use the (somewhat complicated) formula for the probability density function *pdf.*

If, as is the case in this study, there are 5 modes for the female stimulus (vocalization), and each has been rated twice, we have 10 modes in the domain triangle (as defined above). We use ML (maximum likelihood) to estimate the Dirichlet distribution over these 10 modes in order to obtain the mode for female raters for male stimulus for two rating tasks.

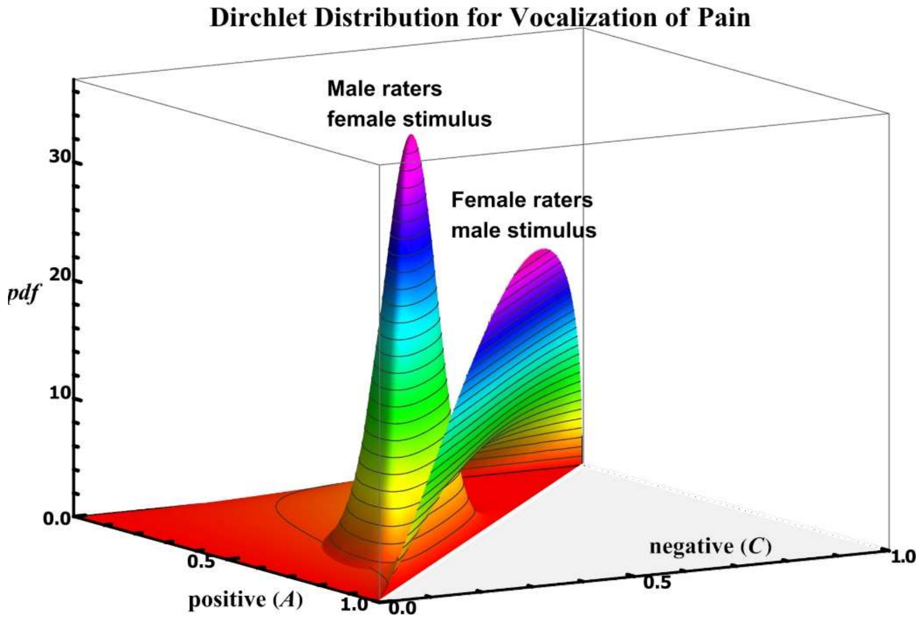


Figure A-1 The *pdf*s of two Dirichlet distributions, one of which has no mode within the domain. Note that the *pdf* of one of the Dirichlet distributions approaches infinity beyond the boundary of the domain (the hypotenuse in this example). Both Dirichlet distributions were determined using ML (maximum likelihood) of the modes of 10 stimuli (either 5 females or 5 males, each rated by the raters twice). Further implications of the (mathematical) divergence of the *pdf* of one of the Dirichlet distributions are discussed in the text. The surface of one *pdf* can be seen inside the other *pdf* surface at low likelihood levels. This visibility is intentional, in order to aid in reading the graph.

The ML method does not ensure, of course, that a mode will exist within the (triangular) domain. It can — and does — happen, that there exists no such mode (Figure A-1). The *pdf* of the ML-estimated Dirichlet distribution may diverge along or beyond one of the boundaries of the domain. In such a case, the coordinates for the mode are undefined and an interpretation of the statistical properties of the ratings become subtle. Such interpretations are to be found in the text for the case of pain vocalizations.

Bayesian estimation of guessing

Each stimulus is rated as exhibiting one of the five facial expressions. We do not expect, but do postulate — as a test — that the facial expression smile (for example) will be rated positive, while the facial expression pain will be rated negative. We use a Bayesian approach to determine the maximum likelihood of a correct probability(!). For each stimulus of each facial expression rated by the females (say), let n_1 be the number of ratings that agree with the postulated rating, while n_2 is the number of ratings that disagree with the postulated rating (then $n_1 + n_2 = n$; $n = 526$ for female raters; $n = 376$ for male raters). In Bayesian statistics, in which the probability s is a random variable, the likelihood function, for this situation, $pdf(s) = \mathcal{L}(s)$ of s is a Beta Distribution

$$\mathcal{L}(\mathcal{B}e(\alpha,\beta),s) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1}(1-s)^{\beta-1} = \frac{\Gamma(n_1+n_2+2)}{\Gamma(n_1+1)\Gamma(n_2+1)} s^{n_1}(1-s)^{n_2}$$

The probability (in Bayesian statistics) of observing a result disagreeing with the postulate is then,

$$\int_0^{\frac{1}{2}} \mathcal{L}(\mathcal{B}e(\alpha,\beta),s)\, ds$$

The most likely probability s_{ML} is the mode. $s_{\mathrm{ML}} = \mathrm{mode} = \frac{\alpha-1}{(\alpha-1)+(\beta-1)}$. We note that the postulate is always s, even if the postulated rating is negative (as in the case of pain).

Paper to-be-signed

Prof. Hermann Prossinger, Ph.D.
Department of Anthropology
University of Vienna
hermann.prossinger@univie.ac.at
consultant

As consultant I hereby declare that my doctoral student, Mgr. Silvia Boschetti, participated on creation of following publications and confirm for each of the publications her proportion of work. To do so I will use CRediT (Contributor Roles Taxonomy; https://credit.niso.org/).

Binter, J., Boschetti, S., Hladký, T., Prossinger, H., Wells, T. J., Jílková, J., & Říha, D. (2022a). Quantifying the Rating Performance of Ambiguous and Unambiguous Facial Expression Perceptions Under Conditions of Stress by Using Wearable Sensors. *In International Conference on Human-Computer Interaction* (pp. 519-529). Springer, Cham. doi: 10.1007/978-3-031-17615-9_36

Conceptualization, Writing Original Draft, Writing — Review and Editing

Binter, J., Boschetti, S., Hladký, T., Prossinger, H., Wells, T. J., Jílková, J., & Říha, D. (2022b). A 'Serious Games' Approach to Decisions of Environmental Impact of Energy Transformation. *In International Conference on Human-Computer Interaction* (pp. 487-495). Springer, Cham. doi: 10.1007/978-3-031-05637-6_31

Conceptualization, Writing Original Draft, Writing — Review and Editing

Binter, J., Prossinger, H., Hladký, T., & Boschetti, S. (In review 2023b). " Wow!" and" Aah": Bimodal Stimuli of Affective States with High and Low Intensity Displays. *Scientific Reports*.

Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Resources, Writing Original Draft, Writing — Review and Editing

Boschetti, S., Prossinger, H., Hladký, T., Machová, K., & Binter, J. (2022). "Eye can't see the difference": Facial Expressions of Pain, Pleasure, and Fear Are Consistently Rated Due to Chance. *Human Ethology*, Volume 37, 046-072. doi:

10.22330/he/37/046-072

Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Writing Original Draft, Writing — Review and Editing

Boschetti, S., Binter, J. and Prossinger, H. (2023a, In Press) Apoidolia: A New Psychological Phenomenon Detected by Pattern Creation with Image Processing Software together with Dirichlet Distributions and Confusion Matrices. *Proceedings in Adaptation, Learning and Optimization*.

Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing Original Draft, Writing — Review and Editing

Boschetti, S., Prossinger, H., Hladký, T., Říha, D., Příplatová, L., Kopecký, R. & Binter, J. (2023b, In Press). Are Patterns Game for Our Brain? AI identifies individual Differences in Rationality and Intuition Characteristics of Respondents Attempting to Identify Random and Non-random Patterns. *In International Conference on Human-Computer Interaction*.

Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing Original Draft, Writing — Review and Editing

Boschetti, S., Prossinger, H., Prossinger-Beck, V., Hladký, T., Říha, D. & Binter, J. (2023c, In Press). Never correct: The novel analysis of differing visual (facial expression) and acoustic (vocalization) bimodal displays of the affective states "pain", "pleasure", and "neutral". *In International Conference on Human-Computer Interaction*.

Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Writing Original Draft, Writing — Review and Editing

Prossinger, H., Hladky, T., Binter, J., Boschetti, S., & Riha, D. (2021). Visual Analysis of Emotions Using AI Image-Processing Software: Possible Male/Female Differences between the Emotion Pairs "Neutral"–"Fear" and "Pleasure"–"Pain". *In The 14th PErvasive Technologies Related to Assistive Environments Conference* (pp. 342-346).

Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing Original Draft, Writing — Review and Editing

Prossinger, H., Binter, J., Machova, K., Riha, D., Boschetti, S. (2022a). Machine Learning Detects Pairwise Associations between SOI and BIS/BAS Subscales, making Correlation Analyses Obsolete. *In Human Interaction & Emerging Technologies (IHIET-AI 2022): Artificial Intelligence & Future Applications* . vol 23. AHFE International, USA. doi: 10.54941/ahfe100903

(also used in PhD thesis of Jakub Binter)

Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Writing — Review and Editing

Prossinger, H., Hladký, T., Boschetti, S., Říha, D., & Binter, J. (2022). Determination of "Neutral"–"Pain","Neutral"- "Pleasure", and "Pleasure"–"Pain" Affective State Distances by Using AI Image Analysis of Facial Expressions. *Technologies,* 10(4), 75. doi: 10.3390/technologies10040075

Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing — Review and Editing

Prossinger, H., Boschetti, S., Říha, D., Pacovský, L. & Binter, J. (2023, In Press) From Stone Age to New Age Statistics: How Neural Networks Overcome the Irreproducibility Problems in Choice Based Profile Creation. *In International Conference on Human-Computer Interaction.*

Data curation, Investigation, Project administration, Writing — Review and Editing

Prague Feb. 22, 2023 Prof. Hermann Prossinger, Ph.D.