

# Brain Intelligence and Artificial Intelligence

## 人脑智能与机器智能

Lecture 13 – Emotion in Brain & AI

**Quanying Liu (刘泉影)**

SUSTech, BME department

Email: [liuqy@sustech.edu.cn](mailto.liuqy@sustech.edu.cn)

# Lecture 13 – Emotion in Brain & AI

---

- **What is emotion?**
- **Emotion theory**
  - Basic emotion theory --> Dimensional emotion theory
  - Theory-driven --> Data-driven
- **Limbic system: amygdala & fear**
  - Innate fear
  - Learned fear
- **Other brain areas**
  - Insular & ventromedial prefrontal cortex (vmPFC)
  - Hypothalamus & brainstem (e.g. PAG)
- **Emotional EEG datasets**
- **Emotional AI**

# What is emotion?

---

love, hate, happiness, sadness, fear, anxiety, ...

1. **Physiological responses to certain kinds of stimuli**, occurring more or less **unconsciously** when the brain detects certain challenging situations. These automatic physiological responses occur within both the brain and the body proper.
  - In the brain they involve changes in arousal levels and in cognitive functions such as attention, memory processing, and decision strategy.
  - In the body proper they involve endocrine, autonomic, and musculoskeletal responses.
2. **Feelings** which are **conscious** experiences often (but not always) accompany these bodily responses. In a certain sense, feelings are accounts our brain creates to represent the physiological phenomena generated by the emotional state.



**Affective neuroscience**

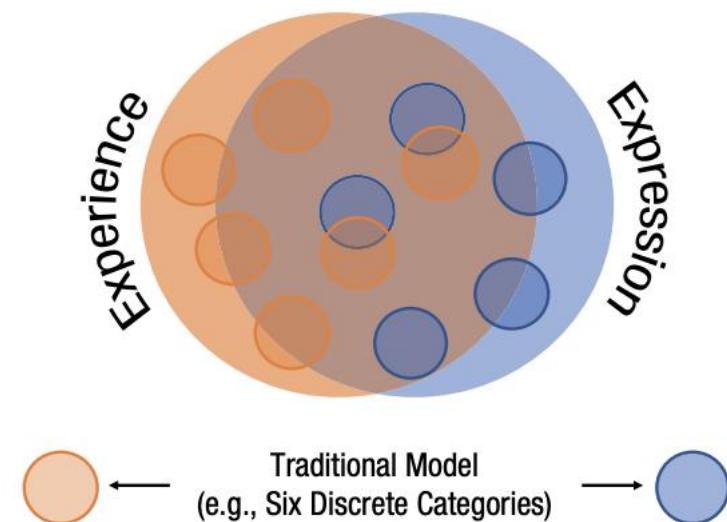
# Is emotion unique to humans?

- **No.** Emotional reactions have been conserved throughout the evolution of species.
- *Behavioral responses* that we typically call **emotional responses** are found in very simple organisms that **may not have consciousness** and thus **not have feelings**.
- For example, a bacterial cell can detect harmful and useful chemicals, and **respond** to these in adaptive ways. Indeed, all organisms must have such capacities to survive and thrive.
- **Emotional responses:** Freezing in danger, Danger avoidance behavior, pleasure seeking, ...
- Distinguishing: emotional **experience** & emotional **expression**

Expressions of anger in animals and humans



These drawings are from Darwin's book, *The Expression of the Emotions in Man and Animals*



# The James–Lange Theory vs The Cannon–Bard Theory

We experience emotion in response to physiological changes in our body.

必须有生理改变才能有情绪。

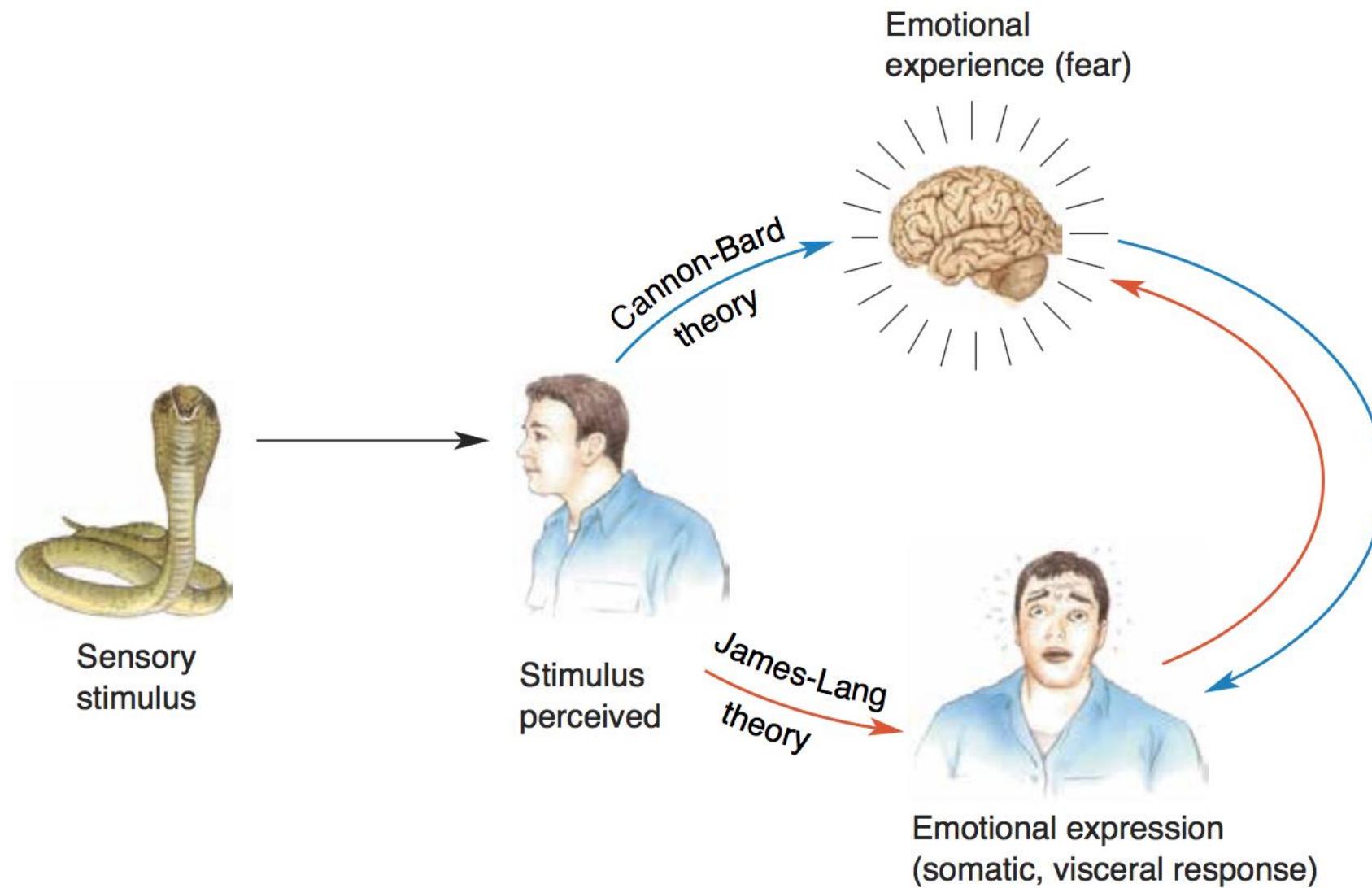
- **情绪源于生理反应**: 情绪的体验是对生理反应的主观感知。我们先经历身体的变化，再将这些变化解释为情绪。
- **刺激 → 生理反应 → 情绪**: 情绪反应的顺序是遇到刺激后，身体首先产生反应，然后我们意识到这种生理变化并体验到对应的情绪。例如，我们看到蛇时先产生颤抖或心跳加快，然后才体验到“恐惧”这一情绪。、
- **情绪的依赖性**: 不同情绪体验基于不同的生理反应。理论认为，若没有生理反应，我们也不会体验到情绪。

Emotions can be experienced even if physiological changes cannot be sensed.

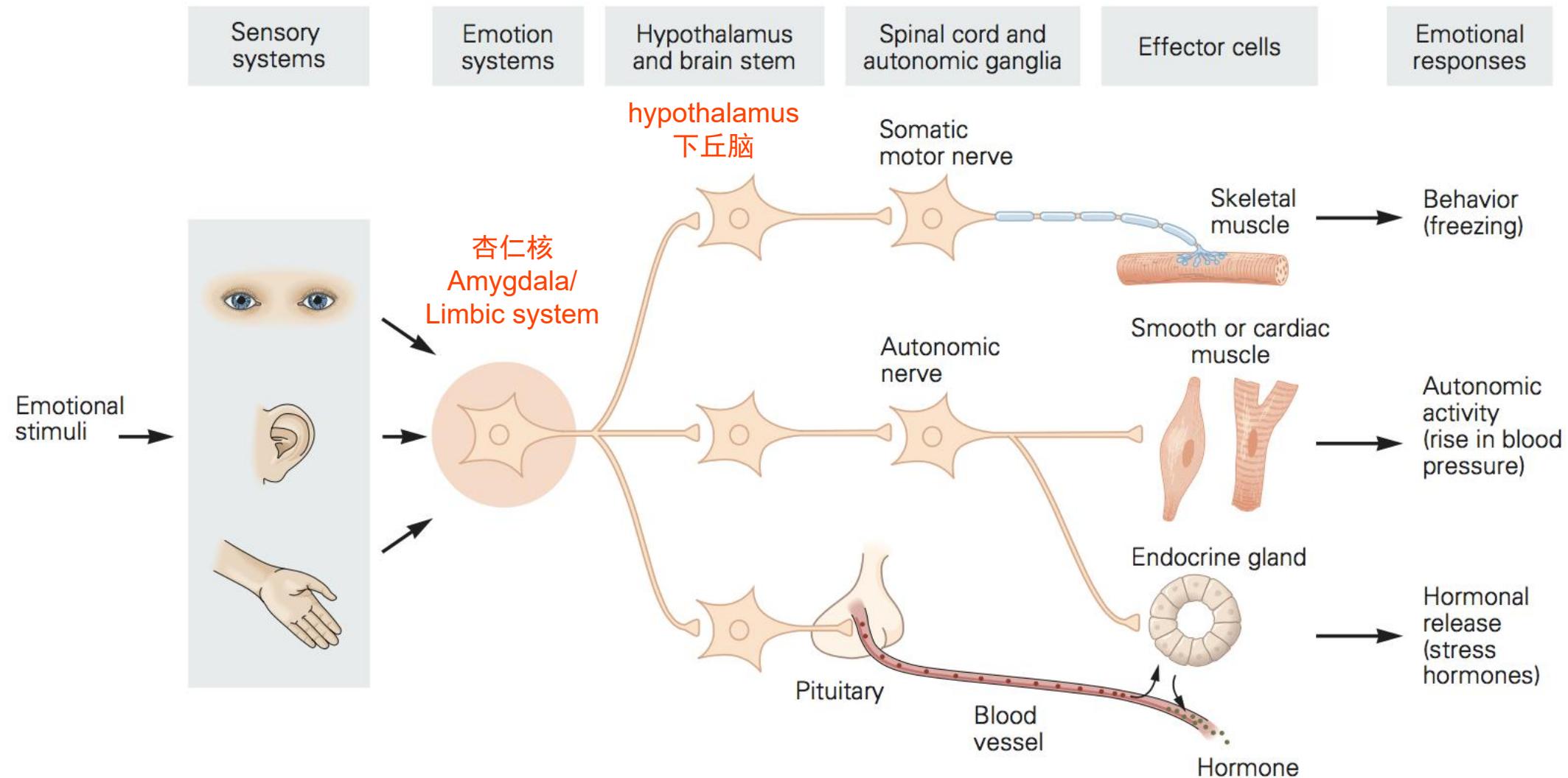
哪怕没有生理反应也能有情绪。

- **情绪和生理反应同时发生**: 当我们面对刺激时，大脑会同时产生情绪体验（如恐惧、愤怒）和对应的身体反应（如心跳加速、出汗）。
- **情绪反应的中心作用**: Cannon 认为丘脑在情绪反应中起到关键作用，接收到刺激后会立即发送信号到大脑皮层以产生情绪，同时传达信息到身体各部分引发生理反应。
- **质疑生理反应的必要性**: 个体可以在没有特定的生理反应的情况下依然体验情绪，这与 James–Lange 理论的因果关系逻辑形成对比。

# The James–Lange Theory vs The Cannon–Bard Theory



# Neural control of emotional responses to external stimuli



# Neural control of emotional responses to external stimuli

Neural control of emotional responses to external stimuli包含哪些神经环路？情绪反应是能被人有意识去控制的吗？

外界刺激引发的情绪反应涉及一系列复杂的神经环路，这些环路在大脑中不同区域之间传递信息，以产生并调节情绪反应。关键的神经环路包括以下几个主要部分：

## 1. 边缘系统：

- 杏仁核 (amygdala)：杏仁核是情绪反应的核心区域，特别是对恐惧、愤怒等负性情绪的处理。当遇到威胁或其他情绪激发的刺激时，杏仁核会迅速被激活，促进应激反应。
- 海马体 (hippocampus)：海马体与记忆密切相关，它通过将情绪与情景记忆联系起来，使人们在未来遇到类似刺激时更容易产生类似的情绪反应。

## 2. 前额叶皮层 (prefrontal cortex, PFC)：

- 腹外侧前额叶皮层 (ventrolateral prefrontal cortex, VLPFC) 和背外侧前额叶皮层 (dorsolateral prefrontal cortex, DLPFC)：这些区域与情绪的认知调节相关，能够帮助我们评估和抑制情绪反应。通过前额叶皮层的认知调节，我们可以更好地控制冲动，做出理性反应。
- 腹内侧前额叶皮层 (ventromedial prefrontal cortex, VMPFC)：VMPFC在整合情绪和决策过程中扮演重要角色，可以帮助抑制杏仁核的活动，从而调节情绪反应的强度。

## 3. 下丘脑-垂体-肾上腺轴 (HPA轴)：

- HPA轴在应激反应中起到关键作用，调节身体对威胁的生理反应（如释放皮质醇），从而产生包括心跳加速、呼吸急促等身体反应，使个体为应对挑战做好准备。

## 4. 腹侧纹状体 (ventral striatum) 和伏隔核 (nucleus accumbens)：

- 这些区域参与奖赏和动机的调节，对愉悦和积极情绪反应尤其重要。伏隔核在愉快情绪的产生中发挥核心作用，与多巴胺系统联系紧密。

## 情绪反应的可控性

情绪反应在一定程度上是可以被有意识控制的，但并非完全。**前额叶皮层**，尤其是DLPFC和VLPFC的调节功能，使得人类可以通过**认知重评、深呼吸**等方式来抑制或重新定义情绪体验。例如，当我们面对压力或不愉快的情绪时，可以通过正念、思考其他事物或更积极的自我暗示来减缓情绪反应。

不过，在强烈的情绪反应或极端应激条件下，如强烈的恐惧或愤怒，**杏仁核**等情绪中心的快速反应可能超出有意识的控制范围，难以立即平复。这种情况下，情绪调节主要依赖于**反复训练**，例如通过**认知行为疗法 (CBT)**、**正念冥想**等，以增强前额叶皮层的调节能力，逐步提高对情绪的控制力。

# Vote

---

## 1. Emotions are universal.

(情绪可以被一些基本的情绪维度来描述和刻画。)

## 2. Emotions are individualized.

(情绪是我个人的。别人无法理解我的情绪。)

# Are Emotions Universal or Individualized?

## Basic emotion theory

There are **universal** expressions and physiological responses for a specific emotion.



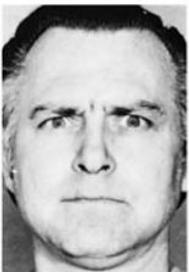
Surprise



Fear



Disgust



Anger



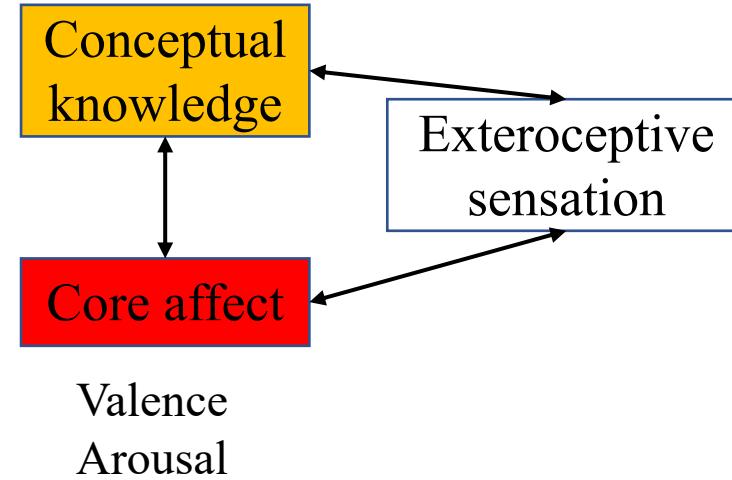
Happiness



Sadness

## The constructivist theory

Emotion is influenced by **personal** experience & culture. It is categorized by **conceptual knowledge** based on core affects(情感) and exteroceptive sensations(感觉).



# Basic emotion theory

“基本情绪理论”（Basic Emotion Theory）将情绪分为具体类型（如快乐、愤怒、恐惧等）。

Certain emotions are thought to be **unique, indivisible** experiences that are **innate and universal** across cultures, an idea that seems a logical extension of Darwin's early observations on the universality of a small number of emotions.

六种情绪：**anger, disgust, fear, happiness, sadness, and surprise.**

Emotion	Description
Astonishment	eyes open, mouth open, eyebrows raised, hands placed over mouth
Contemplation	frown, wrinkle skin under lower eyelids, eyes divergent, head droops, hands to forehead, mouth, or chin, thumb/index finger to lip
Determination	firmly closed mouth, arms folded across breast, shoulders raised
Devotion	face upwards, eyelids upturned, fainting, pupils upwards and inwards, humbling kneeling posture, hands upturned
Happiness	eyes sparkle, skin under eyes wrinkled, mouth drawn back at corners
High spirits, Cheerfulness	smile, body erect, head upright, eyes open, eye brows raised, eyelids raised, nostrils raised, eating gestures (rubbing belly), air suck, lip smacks
Joy	muscle tremble, purposeless movements, laughter, clapping hands, jumping, dancing about, stamping, chuckle/giggle, smile, muscle around eyes contracted, upper lip raised.
Laughter	tears, deep inspiration, contraction of chest, shaking of body, head nods to and fro, lower jaw quivers up/down, lip corners drawn backwards, head thrown backward, shakes, head face red, muscle around eyes contracted, lip press/bite.
Love	beaming eyes, smiling cheeks (when seeing old friend), touch, gentle smile, protruding lips (in chimps), kissing, nose rubs.
Maternal love	touch, gentle smile, tender eyes
Pride	head, body erect, look down on others
Tender (sympathy)	tears

Table 1: Darwin's descriptions of the expressive behavior of positive emotions

# ‘Common view’ of emotion and violations

Predicted by “Common View”							Treated as Nondiagnostic		
Example Event Designated by Experimenter	Insulted by Confederate	Viewing Film of Gore	Asked to Give a Public Speech	Viewing Beautiful Photos	Viewing Film of Funeral	Sudden Loud Noise	Viewing Beautiful Photos	Asked to Give a Public Speech	Insulted by Confederate
Reported Experience (Basic Six)	“Anger”	“Disgust”	“Fear”	“Happiness”	“Sadness”	“Surprise”	“Happiness”	“Fear”	“Anger”
Facial Expression Elicited									
Perceived Emotion (Basic Six)	“Anger”	“Disgust”	“Fear”	“Happiness”	“Sadness”	“Surprise”	“Surprise”	“Happiness”	“Disgust”
						More Plausible Alternative	Awe	Embarrassment	Contempt

**Fig. 1.** Barrett and colleagues' portrayal of the “common view” of emotion and example violations of a “common view” model of emotion. In Barrett and colleagues' portrayal of lay and scientific views on emotion, particular emotion antecedents consistently elicit experiences that are captured by six coarse, mutually exclusive categories—“anger,” “disgust,” “fear,” “happiness,” “sadness,” and “surprise” (a). These

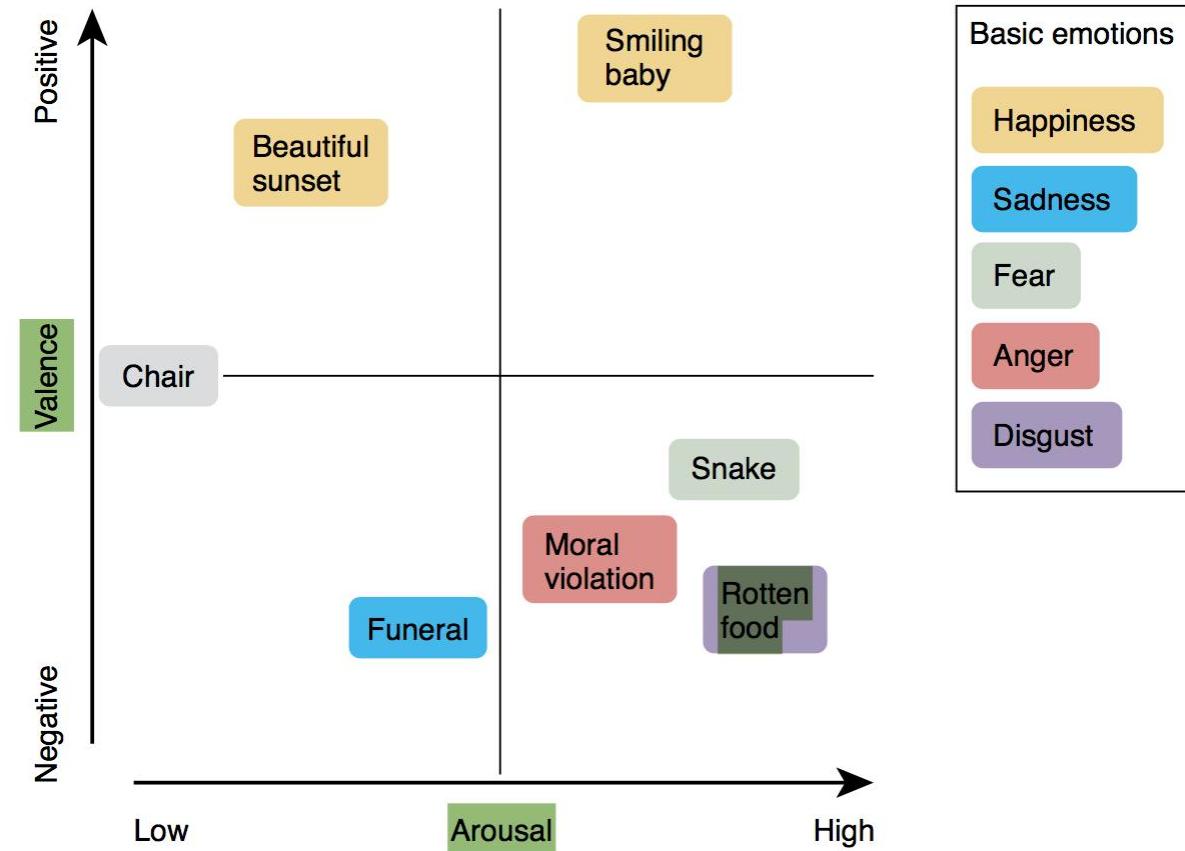
# Dimensional emotion theory 情绪维度理论

**情绪维度理论**(Dimensional Emotion Theory)认为情绪可以通过几个连续的核心维度来量化和描述，而不是将情绪视为孤立的、离散的类别。

“基本情绪理论”(Basic Emotion Theory)将情绪分为具体类型（如快乐、愤怒、恐惧等），而“**情绪维度理论**”认为情绪体验可以在一个或多个维度上变化，并且情绪间并非截然分明。

常见的情绪维度包括唤醒(arousal)、效价(valence)和支配性(dominance)。

- **valence** (“pleasant–unpleasant”)
- **arousal** (“weak emotion–strong emotion”)



那么，情绪到底有多少个维度呢？

# Is emotion high-dimensional?



Alan S. Cowen

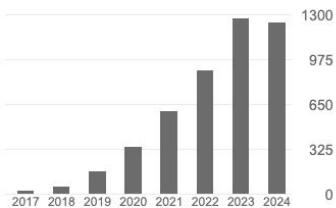
University of California, Berkeley  
Verified email at berkeley.edu - [Homepage](#)

emotion nonverbal behavior affective computing empathic ai ai alignment

FOLLOW

Cited by

	All	Since 2019
Citations	4724	4584
h-index	25	25
i10-index	34	34



TITLE CITED BY YEAR

- Self-report captures 27 distinct categories of emotion bridged by continuous gradients** 1019 2017  
AS Cowen, D Keltner  
   
Proceedings of the national academy of sciences 114 (38), E7900-E7909
- GoEmotions: A dataset of fine-grained emotions** 756 2020  
D Demszky, D Movshovitz-Attias, J Ko, A Cowen, G Nemadé, S Ravi  
arXiv preprint arXiv:2005.00547
- Emotional expression: Advances in basic emotion theory** 497 2019  
D Keltner, D Sauter, J Tracy, A Cowen  
   
Journal of nonverbal behavior 43, 133-160
- Mapping 24 emotions conveyed by brief human vocalization.** 239 2018  
AS Cowen, HA Elfenbein, P Laukka, D Keltner  
   
American Psychologist 74 (6), 698-712.
- Sixteen facial expressions occur in similar contexts worldwide** 225 2021  
AS Cowen, D Keltner, F Schröff, B Jou, H Adam, G Prasad  
   
Nature 589 (7841), 251-257
- What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures** 208 2020  
AS Cowen, X Fang, D Sauter, D Keltner  
   
Proceedings of the National Academy of Sciences 117 (4), 1924-1934
- Mapping the passions: Toward a high-dimensional taxonomy of emotional experience and expression** 189 2019  
A Cowen, D Sauter, JL Tracy, D Keltner  
   
Psychological Science in the Public Interest 20 (1), 69-90
- What the face displays: Mapping 28 emotions conveyed by naturalistic expression.** 171 2019  
AS Cowen, D Keltner  
   
American Psychologist
- The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures** 140 2019  
AS Cowen, P Laukka, HA Elfenbein, R Liu, D Keltner  
   
Nature human behaviour 3 (4), 369-382

Public access [VIEW ALL](#)

0 articles	<b>18 articles</b>
not available	available

Based on funding mandates

Co-authors [VIEW ALL](#)

- Dacher Keltner  
University of California Berkeley >
- Panagiotis Tzirakis  
Imperial College London >
- Alice Baird  
University of Augsburg, Hume AI >
- Disa Sauter  
University of Amsterdam >
- Jessica L Tracy  
Professor of Psychology, Univers... >
- Gautam Prasad  
Google >
- Florian Schröff  
Google Inc. >
- Xia Fang  
Zhejiang University >

# Is emotion high-dimensional?



Alan S. Cowen

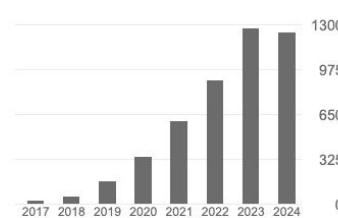
University of California, Berkeley  
Verified email at berkeley.edu - [Homepage](#)

emotion nonverbal behavior affective computing empathic ai ai alignment

[FOLLOW](#)

TITLE	CITED BY	YEAR
Self-report captures 27 distinct categories of emotion bridged by continuous gradients AS Cowen, D Kelchner <small>SCI升级版 综合性期刊1区 SCI基础版 综合性期刊1区 IF 9.4 SWJTU A++ Proceedings of the national academy of sciences 114 (38), E7900-E7909</small>	1019	2017
GoEmotions: A dataset of fine-grained emotions D Demszky, D Movshovitz-Attias, J Ko, A Cowen, G Nemadé, S Ravi <small>arXiv preprint arXiv:2005.00547</small>	756	2020
Emotional expression: Advances in basic emotion theory D Kelchner, D Sauter, J Tracy, A Cowen <small>SCI升级版 心理学3区 IF 1.2 SWUFE B Journal of nonverbal behavior 43, 133-160</small>	497	2019
Mapping 24 emotions conveyed by brief human vocalization. AS Cowen, HA Elfenbein, P Laukka, D Kelchner <small>SCI升级版 心理学1区 IF 12.3 SCI基础版 综合性期刊1区 IF 9.4 SWJTU A++ American Psychologist 74 (6), 698-712.</small>	239	2018
Sixteen facial expressions occur in similar contexts worldwide AS Cowen, D Kelchner, F Schröff, B Jou, H Adam, G Prasad <small>SCI升级版 综合性期刊1区 SCI基础版 综合性期刊1区 IF 50.5 SWJTU A++ Nature 589 (7841), 251-257</small>	225	2021
What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures AS Cowen, X Fang, D Sauter, D Kelchner <small>SCI升级版 综合性期刊1区 SCI基础版 综合性期刊1区 IF 9.4 SWJTU A++ Proceedings of the National Academy of Sciences 117 (4), 1924-1934</small>	208	2020
Mapping the passions: Toward a high-dimensional taxonomy of emotional experience and expression A Cowen, D Sauter, JL Tracy, D Kelchner <small>SCI升级版 心理学1区 IF 18.2 Psychological Science in the Public Interest 20 (1), 69-90</small>	189	2019
What the face displays: Mapping 28 emotions conveyed by naturalistic expression. AS Cowen, D Kelchner <small>SCI升级版 心理学1区 IF 12.3 SWJTU A++ American Psychologist</small>	171	2019
The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures AS Cowen, P Laukka, HA Elfenbein, R Liu, D Kelchner <small>SCI升级版 心理学1区 SCI基础版 医学1区 IF 21.4 Nature human behaviour 3 (4), 369-382</small>	140	2019

Cited by		
	All	Since 2019
Citations	4724	4584
h-index	25	25
i10-index	34	34



Public access	<a href="#">VIEW ALL</a>
0 articles	<a href="#">18 articles</a>
not available	available

Co-authors	<a href="#">VIEW ALL</a>
Dacher Keltner University of California Berkeley	>
Panagiotis Tzirakis Imperial College London	>
Alice Baird University of Augsburg, Hume AI	>
Disa Sauter University of Amsterdam	>
Jessica L Tracy Professor of Psychology, Univers...	>
Gautam Prasad Google	>
Florian Schröff Google Inc.	>
Xia Fang Zhejiang University	>

Sample Text	Label(s)
OMG, yep!!! That is the final answer. Thank you so much!	gratitude, approval
I'm not even sure what it is, why do people hate it	confusion
Guilty of doing this tbph	remorse
This caught me off guard for real. I'm actually off my bed laughing	surprise, amusement
I tried to send this to a friend but [NAME] knocked it away.	disappointment

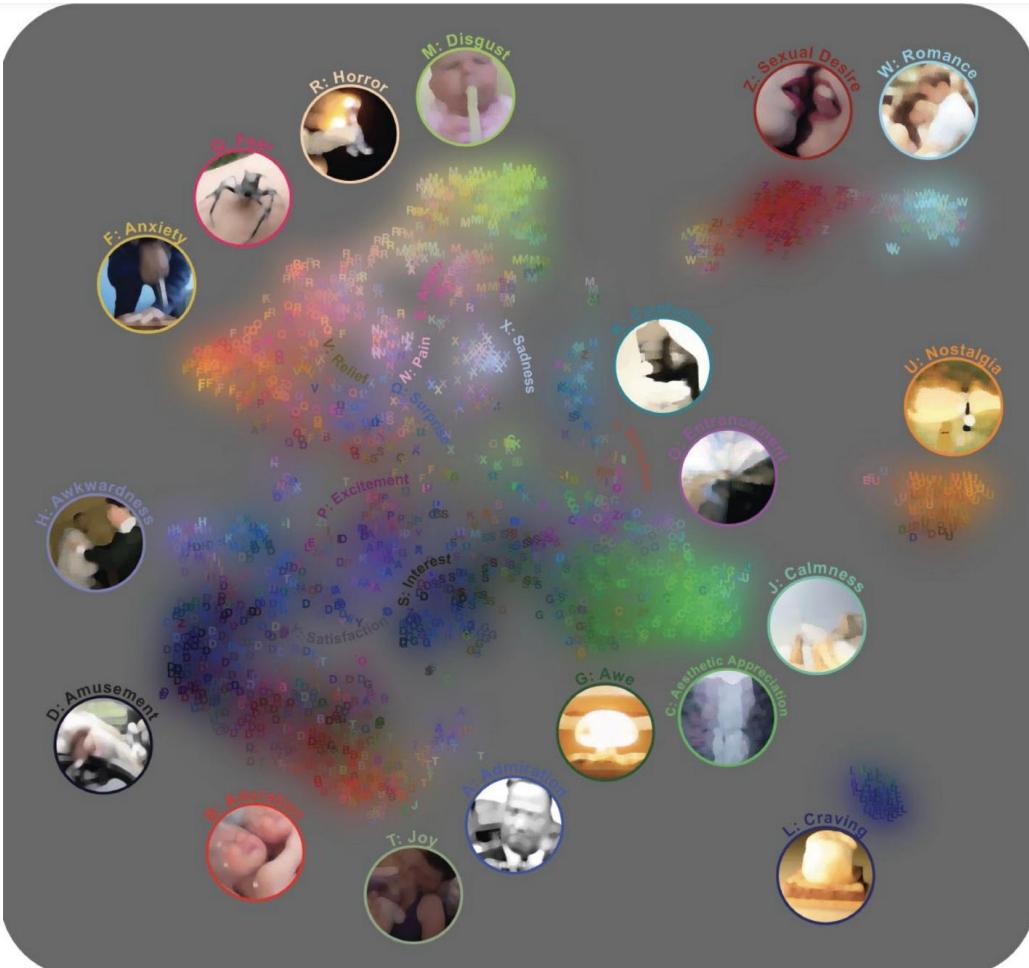
Table 1: Example annotations from our dataset.

<b>Number of examples</b>	58,009
<b>Number of emotions</b>	27 + neutral
<b>Number of unique raters</b>	82
<b>Number of raters / example</b>	3 or 5
<b>Marked unclear or difficult to label</b>	1.6%
<b>Number of labels per example</b>	1: 83% 2: 15% 3: 2% 4+: .2%
<b>Number of examples w/ 2+ raters agreeing on at least 1 label</b>	54,263 (94%)
<b>Number of examples w/ 3+ raters agreeing on at least 1 label</b>	17,763 (31%)

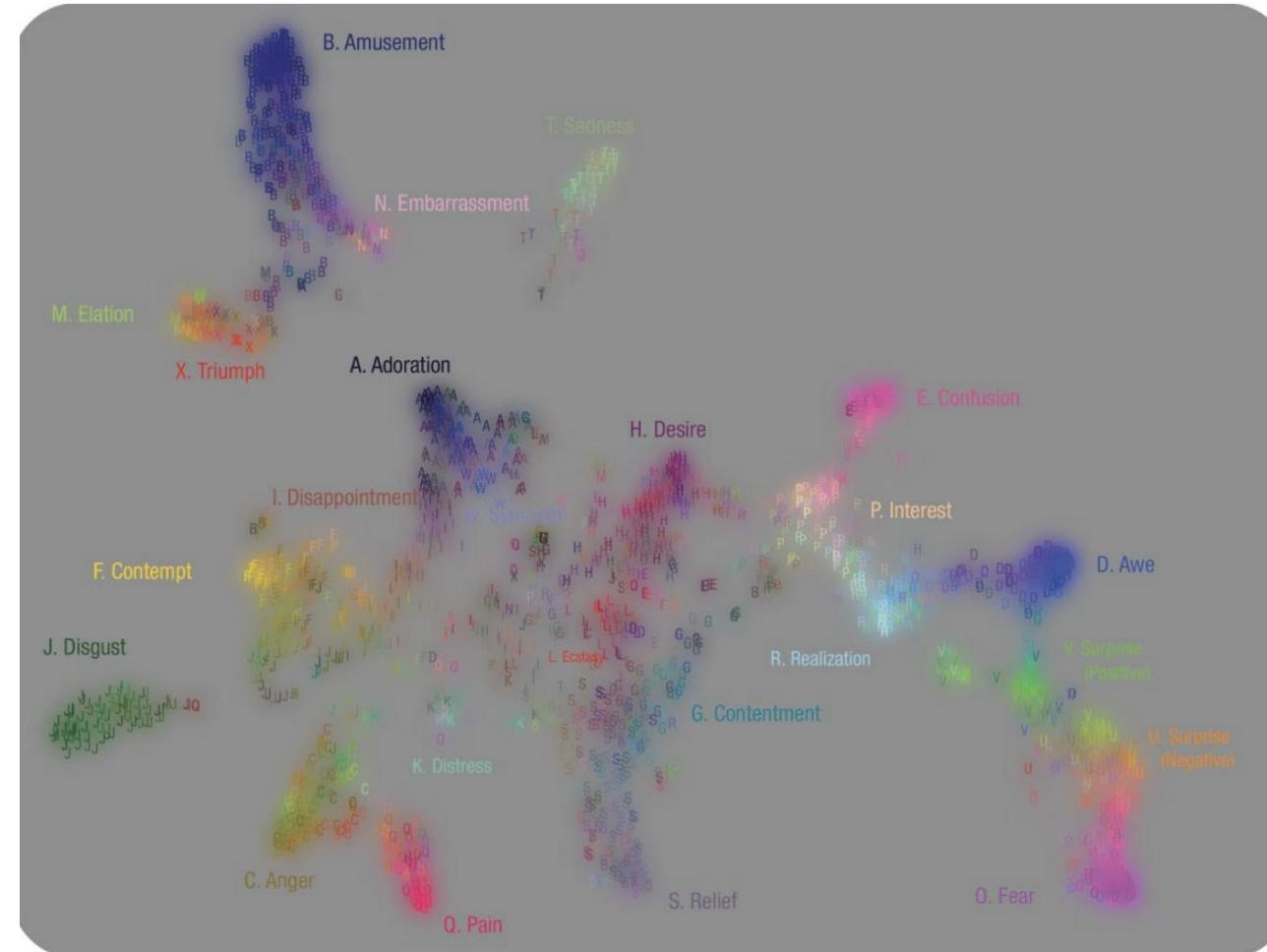
Table 2: Summary statistics of our labeled data.

# Is emotion high-dimensional?

27 varieties of emotional experience  
evoked by 2,185 **videos**

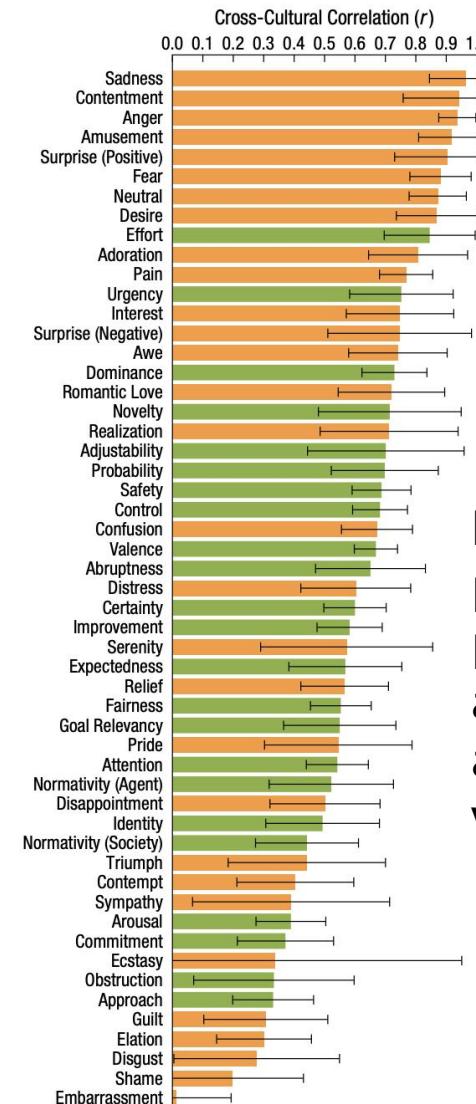
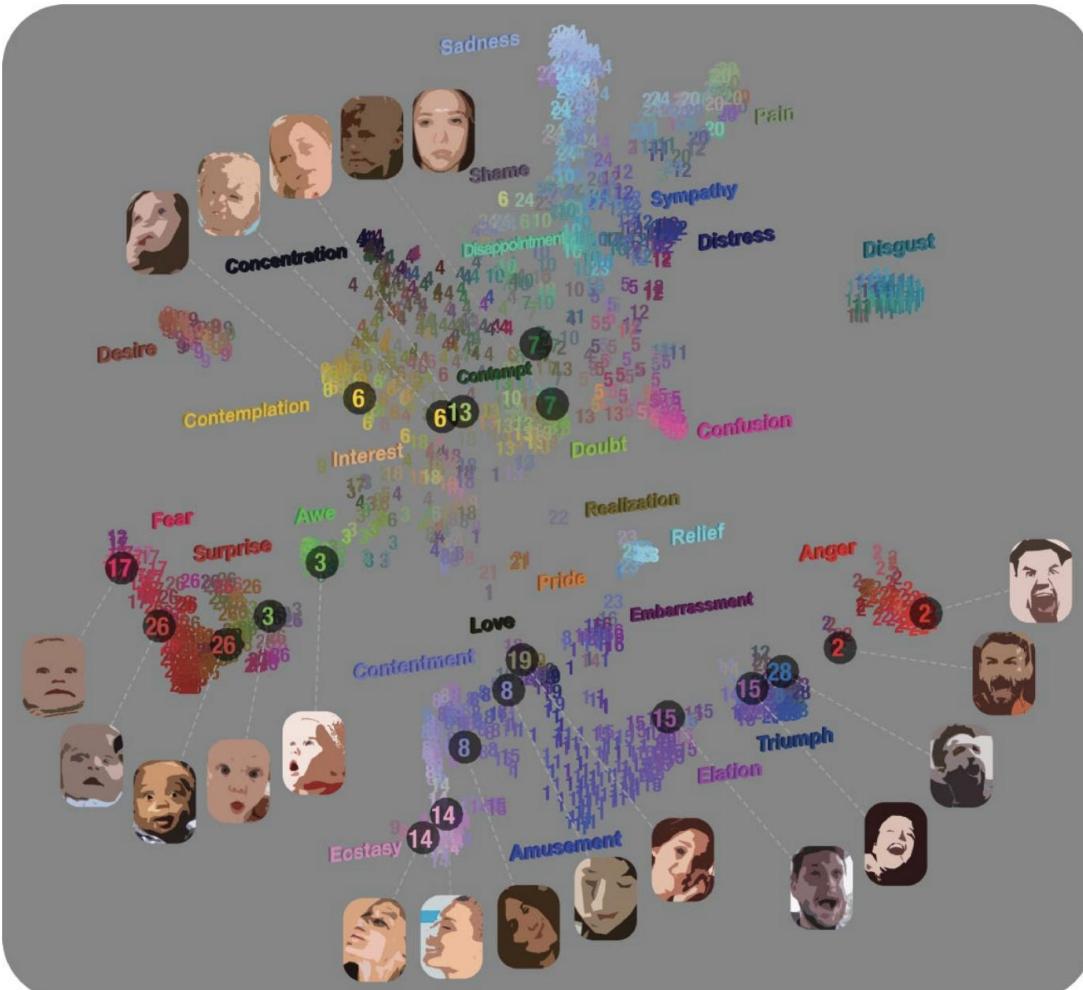


24 varieties of emotion  
recognized in 2,032 **vocal bursts**



# 28-dimensional emotion

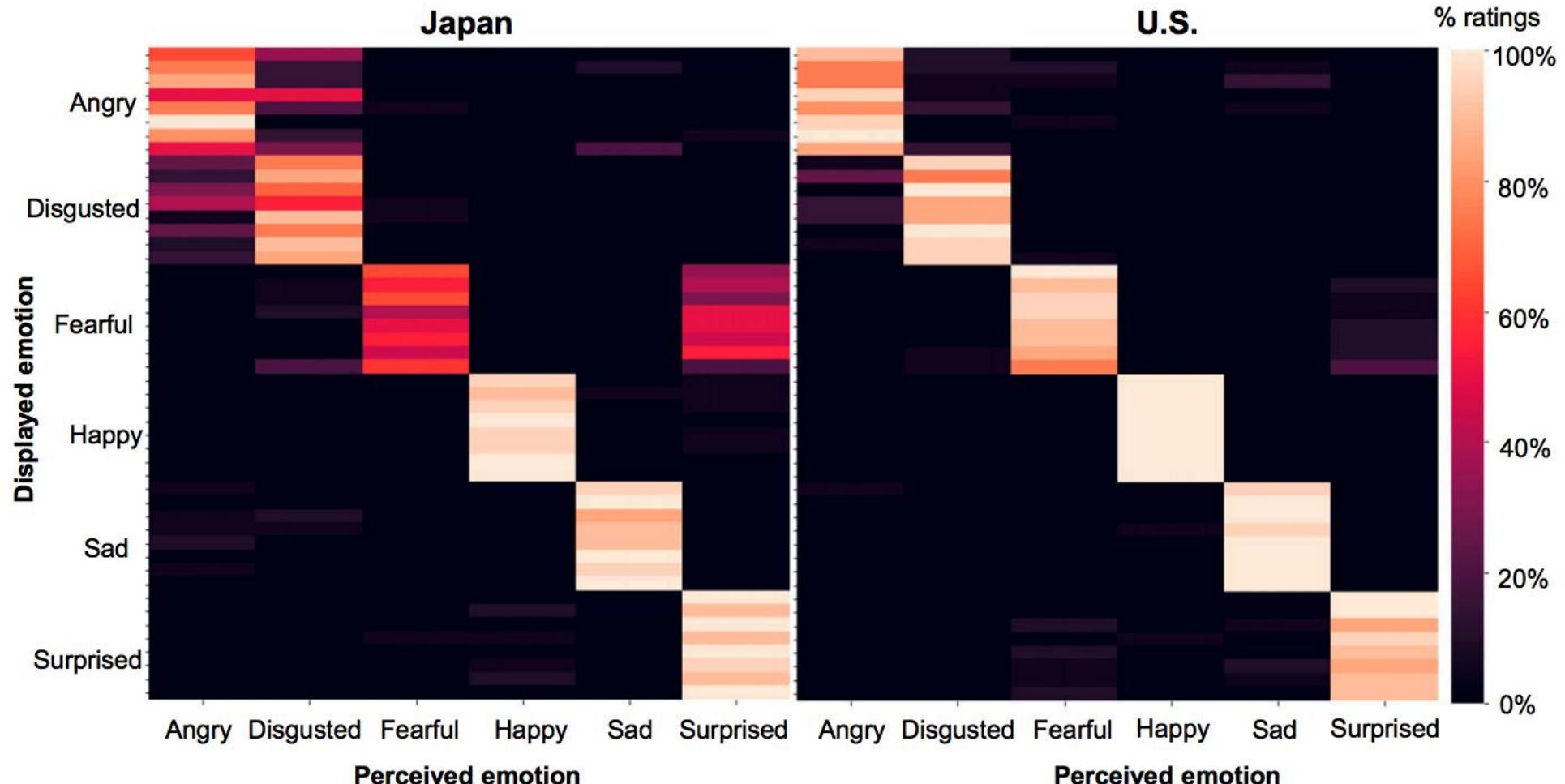
28 varieties of emotion  
recognized in 1,500 facial/bodily expressions



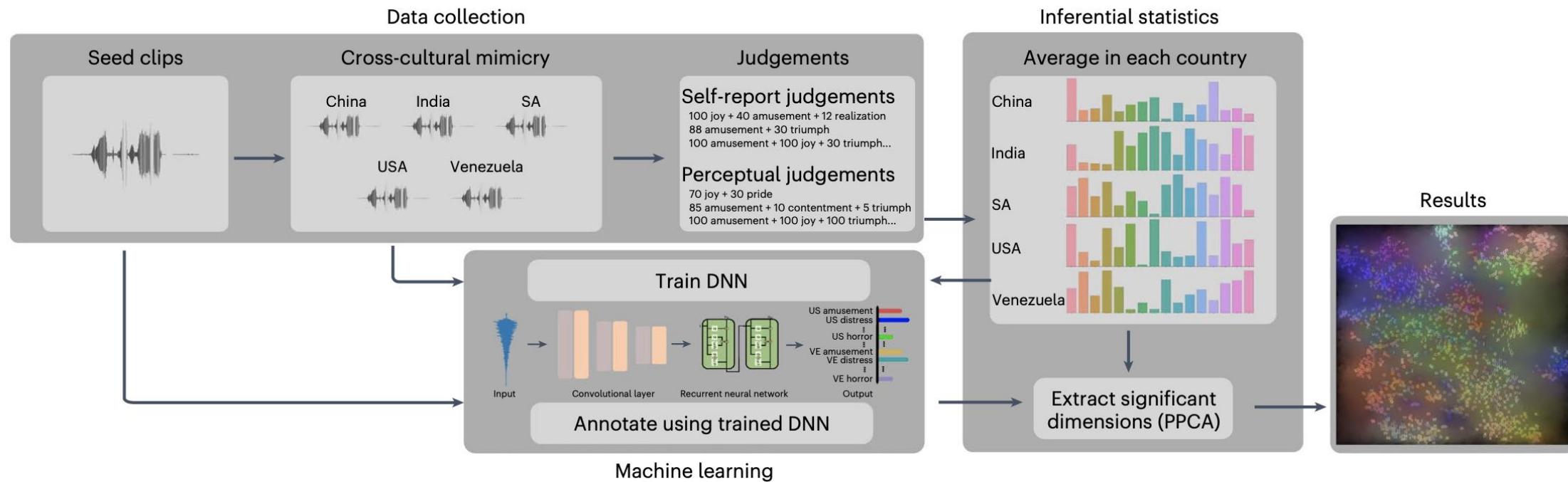
Each judgment is preserved across participants from India and the United States across 2,519 vocalizations.

# Cultural differences in emotion perception

## Explicit categorization task



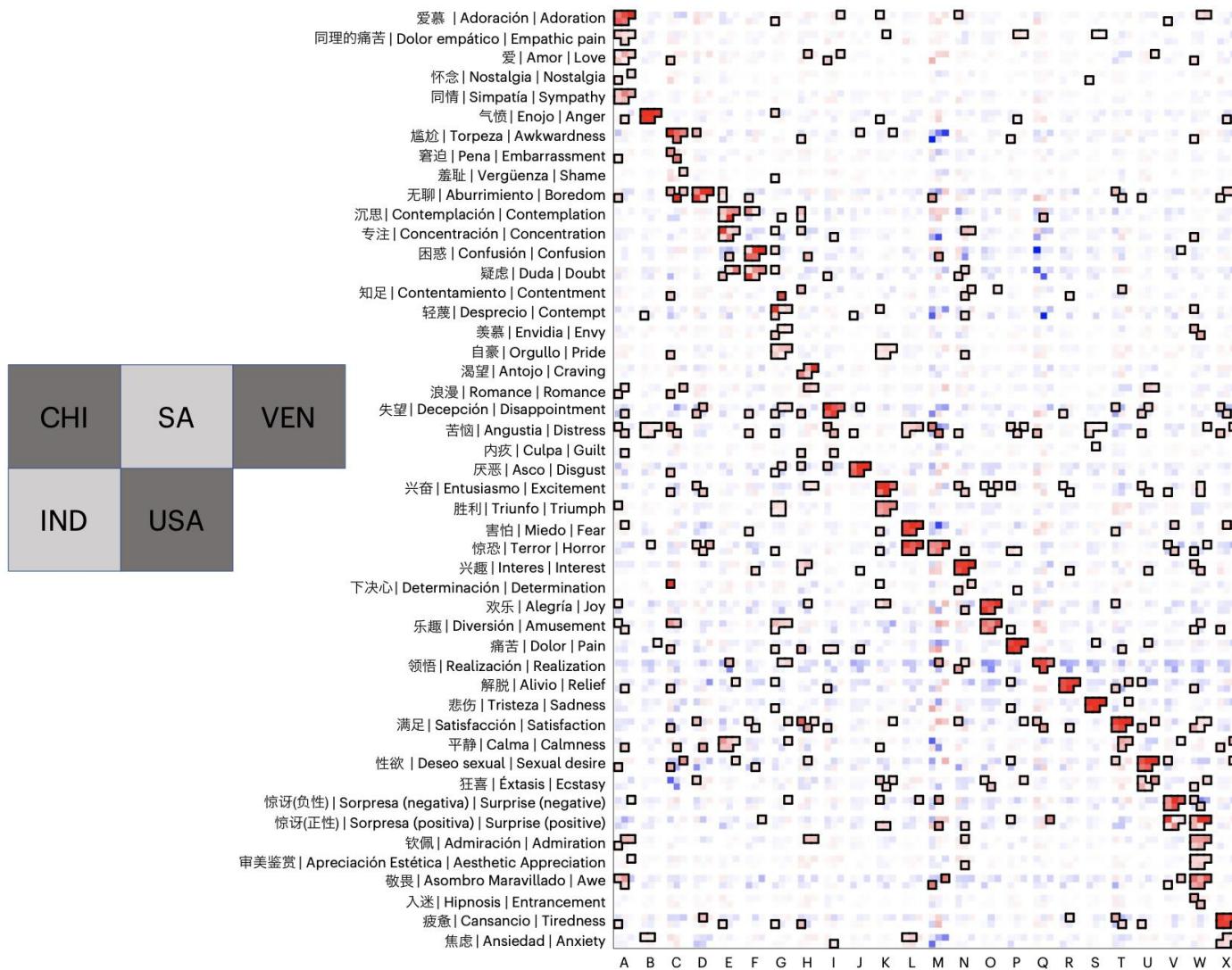
# Cultural differences in emotion expression



**Fig. 1 | Schematic of our experimental and analytic approach.** An initial set of 2,756 seed vocal bursts were rated by 8,941 participants in five countries. These participants also used a computer microphone to record themselves mimicking randomly sampled subsets of up to 30 seed vocal bursts. Thus, each of the resulting 282,906 mimic vocal bursts had a corresponding self-report judgement reflecting what the vocal burst meant to the person making it. An additional 7,879 participants provided perceptual judgements of mimic vocal bursts produced in their own country. We used a DNN to find dimensions of vocal expression that had distinct meanings within or across cultures, independent of demographic and contextual cues, by averaging all self-report and perceptual

ratings corresponding to each seed vocal burst in each country and tasking the DNN with predicting these averages from the mimic vocal bursts (thereby putting a cost on the model predictions being influenced by the speaking voice of the participant forming the imitation; Methods). Finally, we evaluated the DNN on the seed vocal bursts (to which it had no exposure during training) and compared these predictions to the average human judgements of the seed vocal bursts in each country to extract the dimensions of meaning that the DNN successfully identified in distinct vocal modulations (Methods). SA, South Africa; USA, United States of America; VE, Venezuela.

# Cultural differences in emotion expression



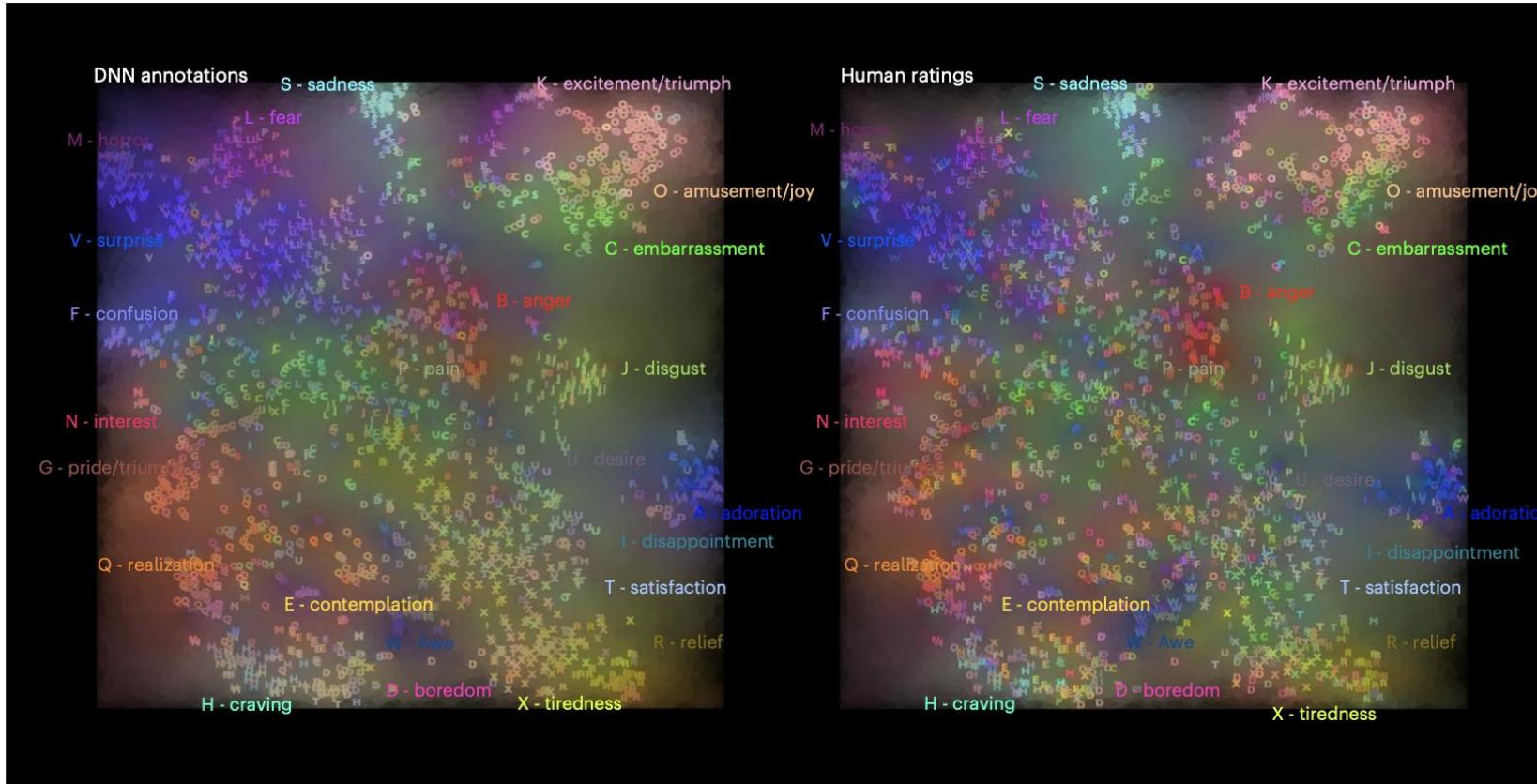
Dimensions of vocal expression that emerged as having distinct meanings within or across cultures

## PPCA on DNN latent representation

### PPCA

PPCA is a dimensionality reduction and statistical modelling technique that decomposes two high-dimensional datasets measuring the same attributes and finds linear combinations of attributes that statistically maximize the covariance explained across datasets<sup>26,27,29,40</sup>. In particular, like more established methods such as partial least-squares correlation analysis (PLSC) and canonical correlation analysis (CCA), PPCA examines the cross-covariance between datasets rather than the variance–covariance matrix within a single dataset. However, whereas PLSC and CCA derive two sets of latent variables,  $\alpha$  and  $\beta$ , maximizing  $\text{cov}(X\alpha_i, Y\beta_i)$  or  $\text{corr}(X\alpha_i, Y\beta_i)$ , PPCA derives only one variable:  $\alpha$ . For an extended validation of the method including a mathematical proof, please refer to ref.<sup>26</sup>.

# Cultural differences in emotion expression



**Fig. 3 | Interactive visualization of vocal bursts along the 24 acoustic dimensions of vocal modulation found to have distinct meanings within or across cultures.** In an interactive visualization (<https://is.gd/iHPOQf>), readers can explore all 282,906 vocal burst mimics, organized by the position of the corresponding seed vocal burst along the 24 dimensions of vocalization we uncovered. We applied t-SNE (Methods) to the concatenated human and DNN annotations of the seed vocal bursts to visualize their distribution along the

24 acoustic dimensions of vocal expression that we found to have distinct shared or culture-specific meanings. Here, colours in the plot on the left, as well as in the interactive visualization, represent DNN annotations projected onto the 24 dimensions. Colours in the plot on the right represent average human intensity ratings projected onto the 24 dimensions. Similarity in colour between the two plots reflects similarities between DNN predictions and human perceptual ratings.

# Hume AI



Alan S. Cowen

University of California, Berkeley  
Verified email at berkeley.edu - [Homepage](#)

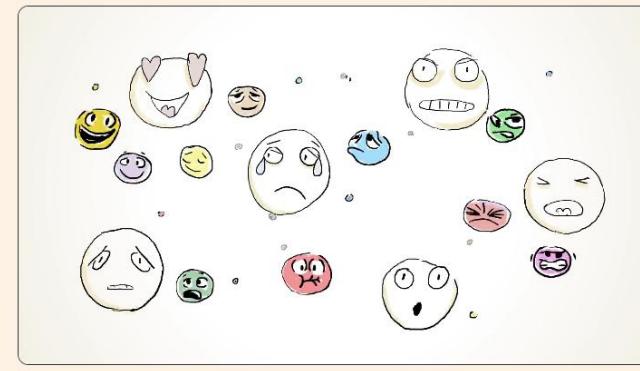
emotion nonverbal behavior affective computing empathic ai ai alignment

FOLLOW

TITLE	CITED BY	YEAR
Emotional palette: a computational mapping of aesthetic experiences evoked by visual art E Stamkou, D Keltner, R Corona, E Aksoy, AS Cowen <small>SCI升级版 综合性期刊2区 SCI基础版 综合性期刊3区 IF 3.8 SWJTU A++ SWUFE B</small> Scientific Reports 14 (1), 19932	5	2024
How emotion is experienced and expressed in multiple cultures: a large-scale experiment across North America, Europe, and Japan AS Cowen, JA Brooks, G Prasad, M Tanaka, Y Kamitani, V Kirilyuk, ... <small>SCI升级版 心理学3区 IF 2.6</small> Frontiers in Psychology 15, 1350631	3	2024
247. Reverse Engineering an Emotion Expression Estimate Classifier for Depressed Mood V Tiruvadi, A Baird, L Schooler, J Brooks, C Gagne, P Tzirakis, H Li, M Oh, ... <small>SCI升级版 医学1区 SCI基础版 医学1区 IF 9.6 SWJTU A++</small> Biological Psychiatry 95 (10), S200-S201		2024
Empathic artificial intelligence systems A Cowen, D Keltner, B Schoenfeld US Patent App. 18/322,347		2024
The NeurIPS 2023 Machine Learning for Audio Workshop: Affective Audio Benchmarks and Novel Data A Baird, R Manzelli, P Tzirakis, C Gagne, H Li, S Allen, S Dieleman, ... arXiv preprint arXiv:2403.14048		2024
Deep learning reveals what facial expressions mean to people in different cultures JA Brooks, L Kim, M Opara, D Keltner, X Fang, M Monroy, R Corona, ... <small>SCI升级版 综合性期刊2区 SCI基础版 综合性期刊2区 IF 4.6</small> Iscience 27 (3)	2	2024
The 6th affective behavior analysis in-the-wild (abaw) competition D Kollias, P Tzirakis, A Cowen, S Zafeiriou, I Kotsia, A Baird, C Gagne, ... Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern ...	54	2024
The muse 2023 multimodal sentiment analysis challenge: Mimicked emotions, cross-cultural humour, and personalisation L Christ, S Amiriparian, A Baird, A Kathan, N Müller, S Klug, C Gagne, ... Proceedings of the 4th on Multimodal Sentiment Analysis Challenge and ...	35	2023
Muse 2023 challenge: Multimodal prediction of mimicked emotions, cross-cultural humour, and personalised recognition of affects S Amiriparian, L Christ, A König, A Cowen, EM Meßner, E Cambria, ... Proceedings of the 31st ACM International Conference on Multimedia, 9723-9725	16	2023
The acm multimedia 2023 computational paralinguistics challenge: Emotion share & requests BW Schuller, A Batliner, S Amiriparian, A Barnhill, M Gerczuk, ... Proceedings of the 31st ACM International Conference on Multimedia, 9635-9639	13	2023

## About Hume

Hume is a research lab and technology company. Our mission is to ensure that artificial intelligence is built to serve human goals and emotional well-being.



## Our vision

We envision a future where AI systems use scientific approaches to fulfill human needs. Emotional intelligence is the missing ingredient needed to build AI systems that proactively find ways to improve your quality of life.

### Hume AI

<https://www.hume.ai> › blog › series-b-evi-announcement

## Hume Raises \$50M Series B and Releases New Empathic ...

Mar 25, 2024 – Hume AI raised a \$50m Series B round led by EQT Ventures and joined by Union Square Ventures, Nat Friedman & Daniel Gross, Metaplanet, Northwell Holdings, ...

### Hume AI

<https://www.hume.ai> › blog › hume-ai-raises-usd12-7m...

## Hume AI Raises \$12.7M in Series A Funding

Feb 16, 2023 – We're pleased to announce that we've raised a \$12.7M Series A led by Union Square Ventures! We're excited to be partnering with leading ...

# From theory-driven to data-driven



TODAY | OUR APPROACH

## The full spectrum of emotion

Hume's scientists are revolutionizing emotion study through data-driven methods, employing computational techniques, and analyzing vast datasets to understand the spectrum of human emotions.

They've gathered millions of reactions across videos, music, and art studied brain mechanisms of emotion, explored ancient sculptures' expressions, and applied deep learning to global video expressions. Their research uncovers over 30 emotion dimensions.



DAVID HUME ..... CHARLES DARWIN ..... THE BASIC 6 ..... VALENCE AND AROUSAL ..... OUR APPROACH

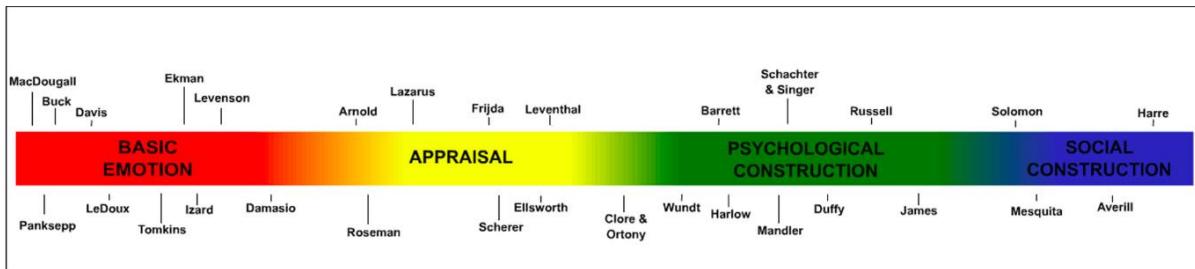


# theory-driven vs data-driven

理论驱动：基于理论设计实验、采集数据、  
进行分析、验证理论

Basic emotion theory 基本情绪理论

Dimensional emotion theory 情绪维度理

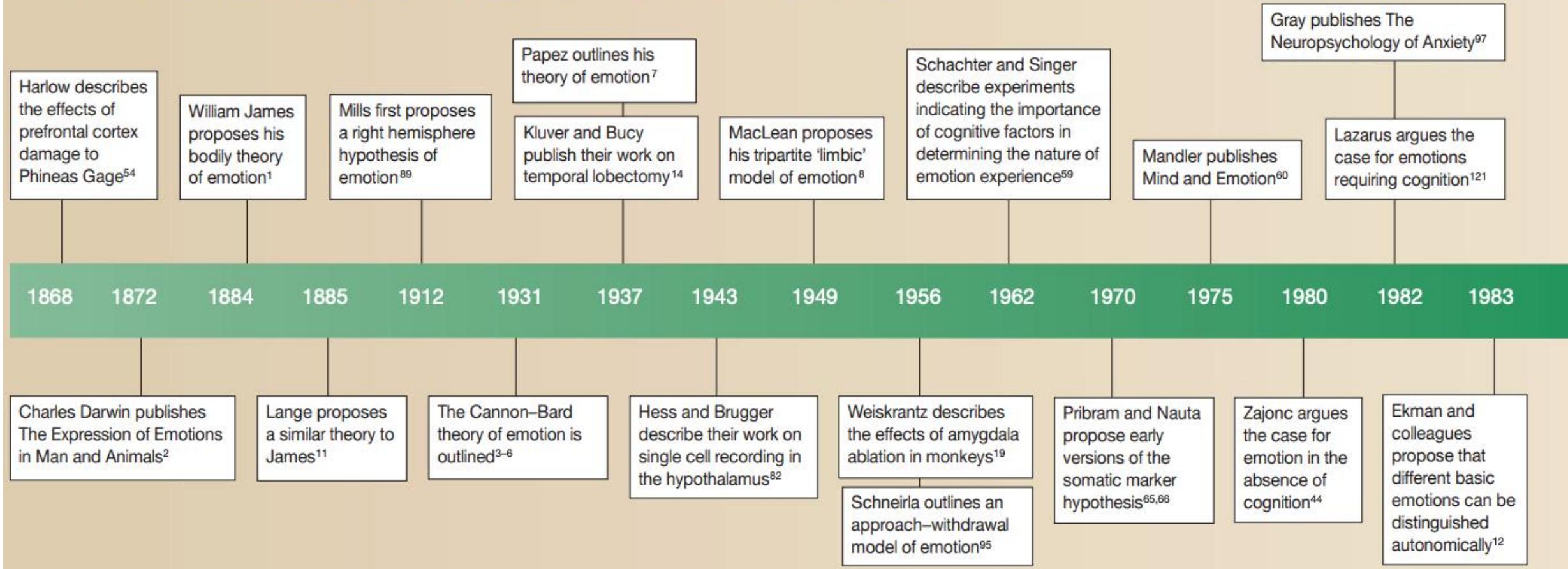


**Figure 1.** Perspectives on emotion can be loosely arranged along a continuum. We have populated this continuum with representative theorists/researchers drawn from the field of psychology. We distinguish four “zones”: (1) basic emotion, in red, e.g., MacDougall (1908/1921), Panksepp (1998), Buck (1999), Davis (1992), LeDoux (2000), Tomkins (1962, 1963), Ekman (1972), Izard (1993), Levenson (1994), and Damasio (1999); (2) appraisal, in yellow, e.g., Arnold (1960a, 1960b), Roseman (1991), Lazarus (1991), Frijda (1986), Scherer (1984), Smith and Ellsworth (1985), Leventhal (1984), and Clore and Ortony (2008); (3) psychological construction, in green, e.g., Wundt (1897/1998), Barrett (2009), Harlow and Stagner (1933), Mandler (1975), Schachter and Singer (1962), Duffy (1941); Russell (2003), and James (1884); (4) social construction, in blue, e.g., Solomon (2003), Mesquita (2010), Averill (1980), and Harré (1986). Given space constraints, as well as the goals of this article, we have limited ourselves to a subset of the many theorists/researchers who might have been included on this continuum (e.g., those who only study one aspect of emotion were not included in this figure).

Gross, James, and Lisa Feldman Barrett. "Emotion generation and emotion regulation: One or two depends on your point of view." *Emotion review*

# Development of Emotion Theory

## Timeline | Historical milestones in understanding the emotional brain



# theory-driven vs data-driven

理论驱动：基于理论设计实验、采集数据、进行分析、验证理论

Basic emotion theory 基本情绪理论

Dimensional emotion theory 情绪维度理

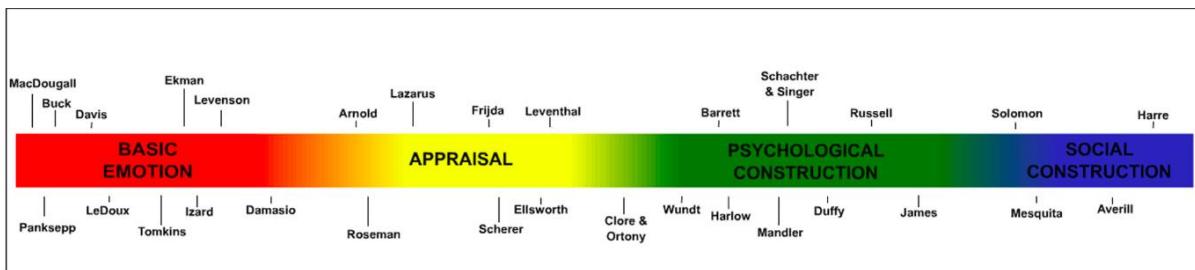


Figure 1. Perspectives on emotion can be loosely arranged along a continuum. We have populated this continuum with representative theorists/researchers drawn from the field of psychology. We distinguish four “zones”: (1) basic emotion, in red, e.g., MacDougall (1908/1921), Panksepp (1998), Buck (1999), Davis (1992), LeDoux (2000), Tomkins (1962, 1963), Ekman (1972), Izard (1993), Levenson (1994), and Damasio (1999); (2) appraisal, in yellow, e.g., Arnold (1960a, 1960b), Roseman (1991), Lazarus (1991), Frijda (1986), Scherer (1984), Smith and Ellsworth (1985), Leventhal (1984), and Clore and Ortony (2008); (3) psychological construction, in green, e.g., Wundt (1897/1998), Barrett (2009), Harlow and Stagner (1933), Mandler (1975), Schachter and Singer (1962), Duffy (1941); Russell (2003), and James (1884); (4) social construction, in blue, e.g., Solomon (2003), Mesquita (2010), Averill (1980), and Harré (1986). Given space constraints, as well as the goals of this article, we have limited ourselves to a subset of the many theorists/researchers who might have been included on this continuum (e.g., those who only study one aspect of emotion were not included in this figure).

Gross, James, and Lisa Feldman Barrett. "Emotion generation and emotion regulation: One or two depends on your point of view." *Emotion review*

数据驱动：从海量数据中进行分析、表示、聚类，从而提出新的情绪理论

Larger-scale, more diverse studies of naturalistic emotional behavior, 对人类情绪表达和情绪体验数据进行分析

关键：海量数据 + AI分析方法

记录人类观看视频、图片、音乐、艺术作品等等时候的感叹声(vocal burst)、脸部表情(facial expression)、对话声(dialogue speech)、文本表达(text)等等数据

Keltner, Dacher, Jeffrey A. Brooks, and Alan Cowen. "Semantic space theory: Data-driven insights into basic emotions." *Current Directions in Psychological Science* (2023)

## 情绪相关的脑区在哪里？

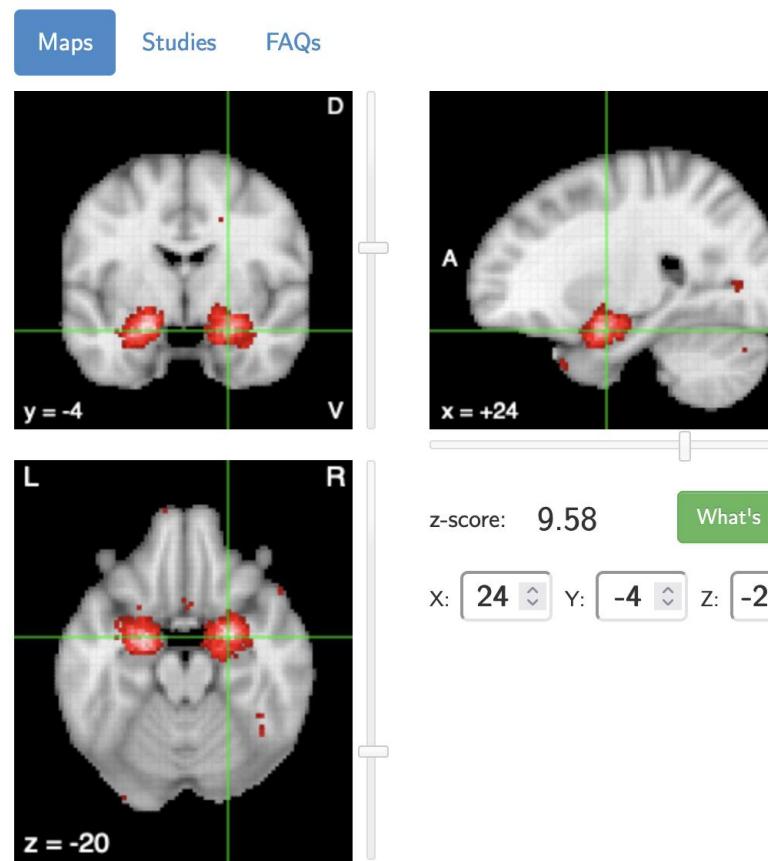
特定的脑区？

很多脑区分布式协同表示？

# Brain regions to represent emotions

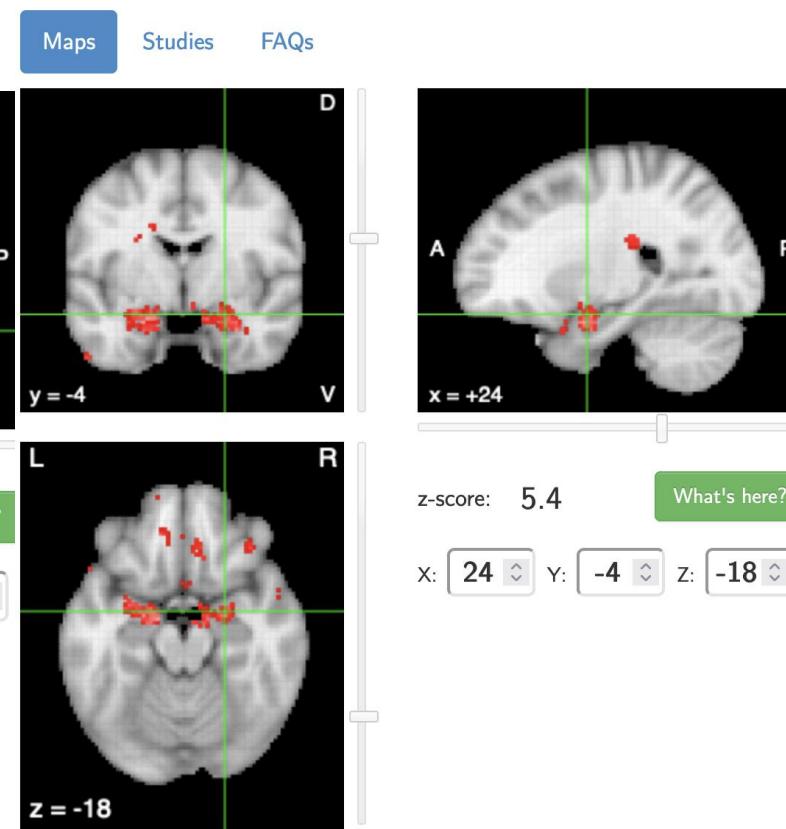
happy

An automated meta-analysis of 225 studies



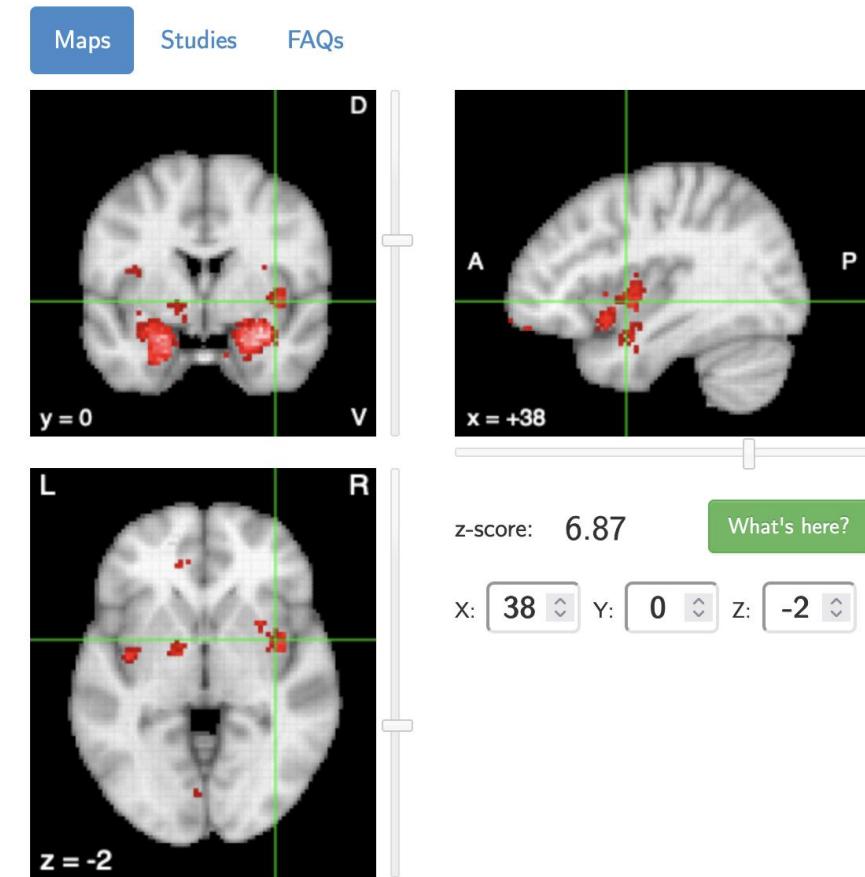
anger

An automated meta-analysis of 89 studies

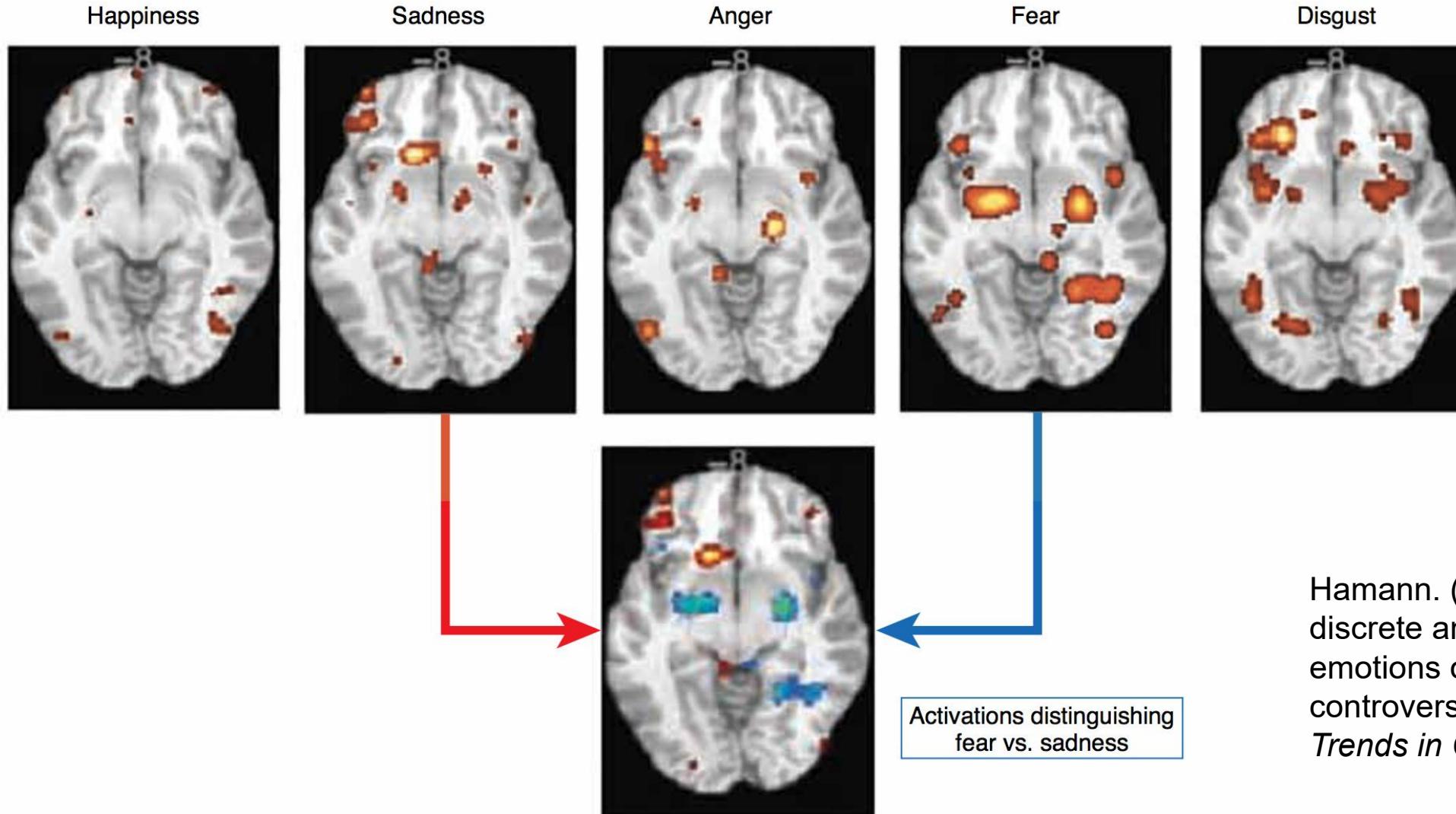


disgust

An automated meta-analysis of 103 studies



# Brain activation associated with 5 basic emotions



Hamann. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in Cognitive Sciences*

## 情绪是在特定的脑区编码，还是分布在很多脑区协同编码？



情绪的编码通常涉及多个脑区的协同作用，而不是仅仅局限于某个特定的脑区。研究表明，情绪处理是一个复杂的过程，涉及大脑多个区域的互动。以下是一些在情绪处理过程中起重要作用的脑区：

1. **杏仁核**：杏仁核是情绪调节和情绪感知的核心区域，尤其在恐惧和焦虑等情绪反应中起着关键作用。
2. **前额叶皮层**：前额叶皮层与情绪的调节、抑制以及决策相关，它在情绪的评估和调控中发挥重要作用。
3. **海马体**：海马体在情绪记忆的形成和回忆中起到重要作用，特别是在情绪相关的记忆处理时。
4. **岛叶**：岛叶与情绪的生理反应和意识体验密切相关，尤其在情绪的感受层面。
5. **扣带回**：扣带回与情绪反应、决策和情绪的认知评估有关。

情绪的产生并非单一脑区的活动，而是多个脑区之间的协同作用。这些脑区的功能不仅是局部的，还可能通过脑网络的相互连接进行调控，例如前额叶皮层通过其下游影响杏仁核的反应，从而调节情绪反应。

因此，情绪的编码是大脑多个区域合作的结果，体现了脑区间的动态交互与协同工作。

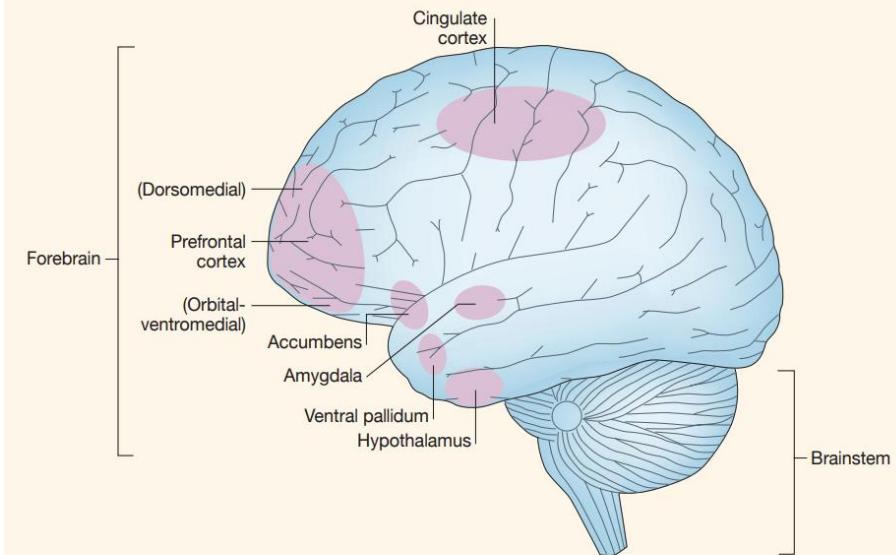
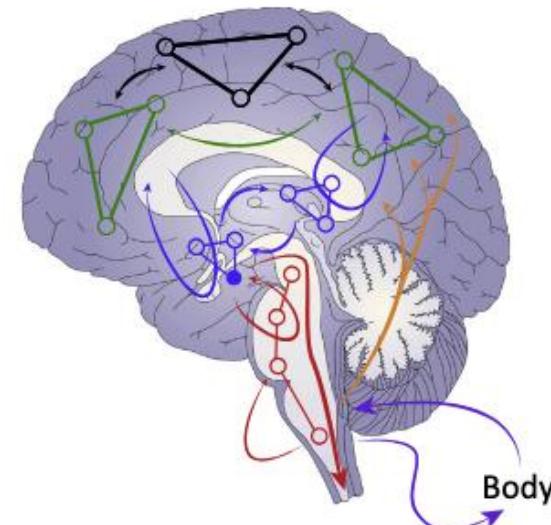


Figure 4 | Key structures within a generalized emotional brain. The figure does not show the relative depths of the various structures, merely their two-dimensional location within the brain schematic. As this is a lateral view, only one member of bilateral pairs of structures can be seen. Anatomical image adapted, with permission, from REF. 123 © (1996) Appleton & Lange.

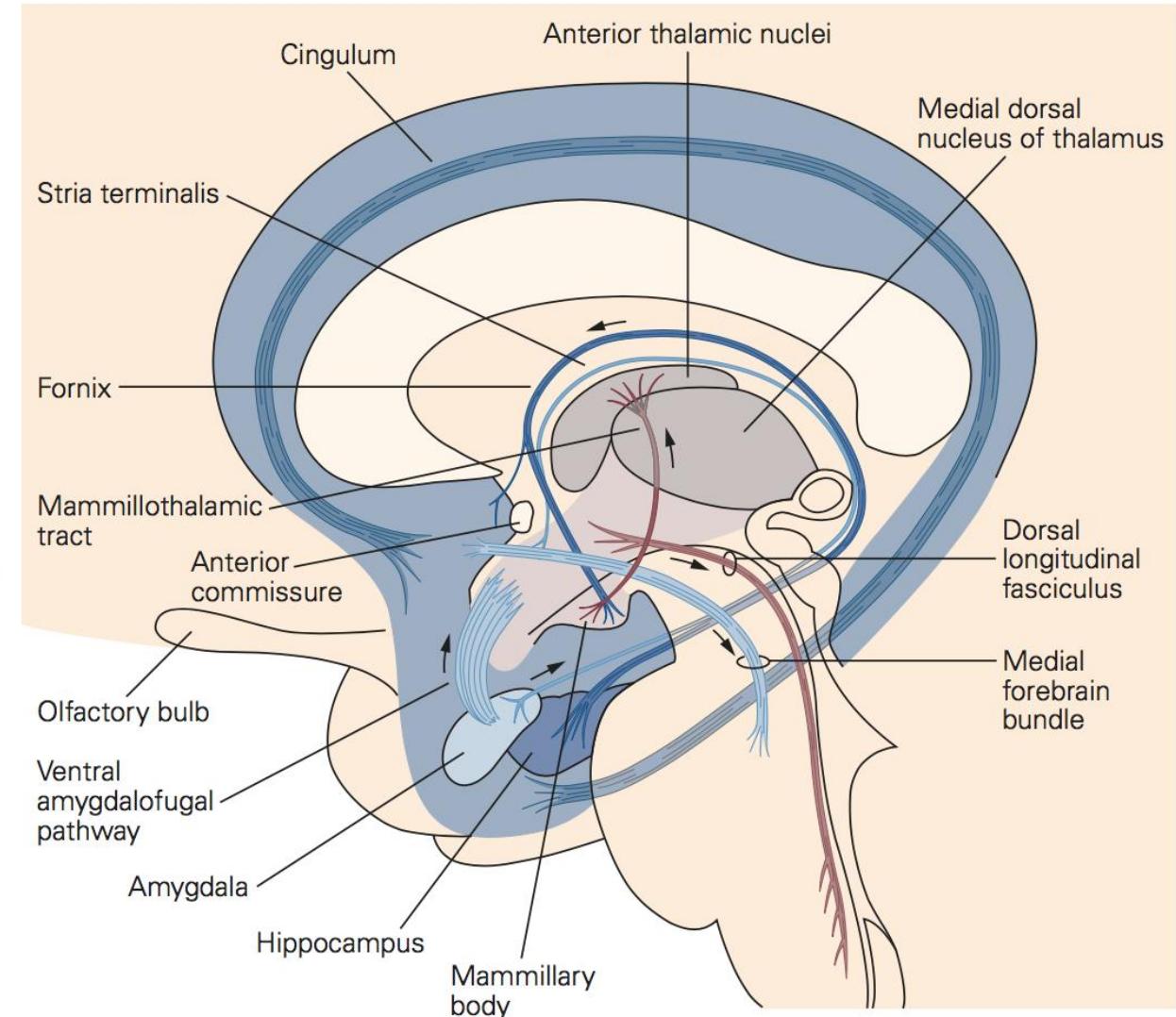
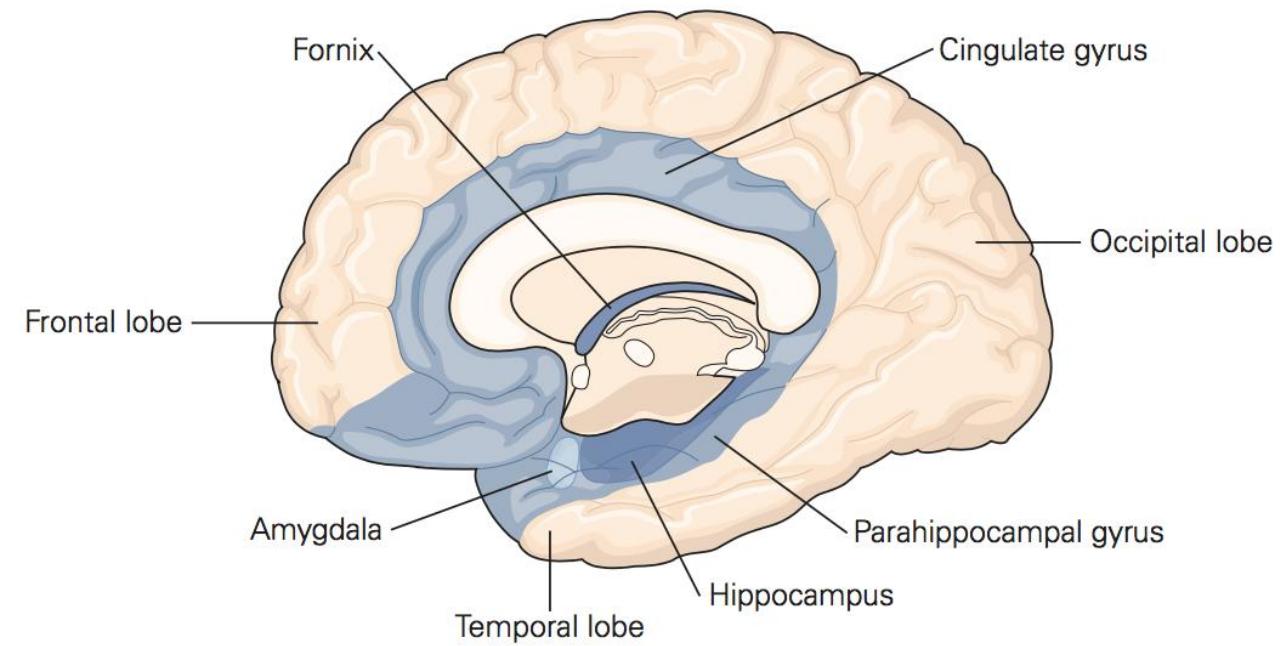
Dalgleish, Tim. "The emotional brain." *Nature Rev. Neurosci.* (2004).



Pessoa, Luiz. "A network model of the emotional brain." *Trends in cognitive sciences* (2008).

# Limbic system 边缘系统

The limbic system consists of the limbic lobe and deep-lying structures.

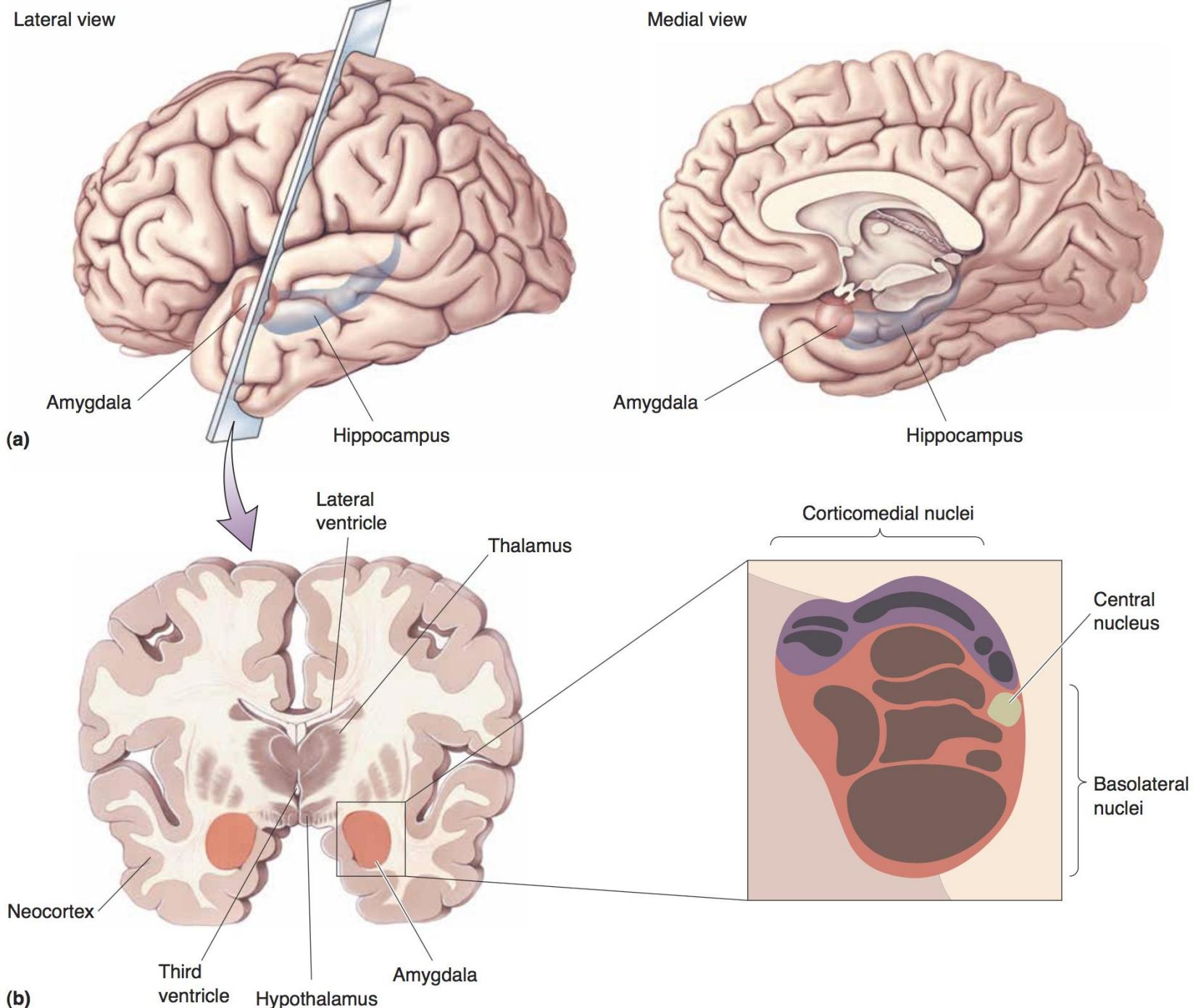


# Amygdala

## A cross section of the amygdala

(a) Lateral and medial views of the temporal lobe, showing the location of the amygdala in relation to the hippocampus.

(b) The brain is sectioned coronally to show the amygdala in cross section. **The basolateral nuclei (surrounded by red)** receive visual, auditory, gustatory, and tactile afferents. The **corticomedial nuclei (surrounded by purple)** receive olfactory afferents.



# Amygdala and induced fear

Brain activity in response to fearful faces showed more amygdala activity than in response to faces with neutral expressions (right figure).

Patients with damage to the amygdala also **fail** to recognize facial expressions of fear and do **not** generate autonomic fear responses to these.

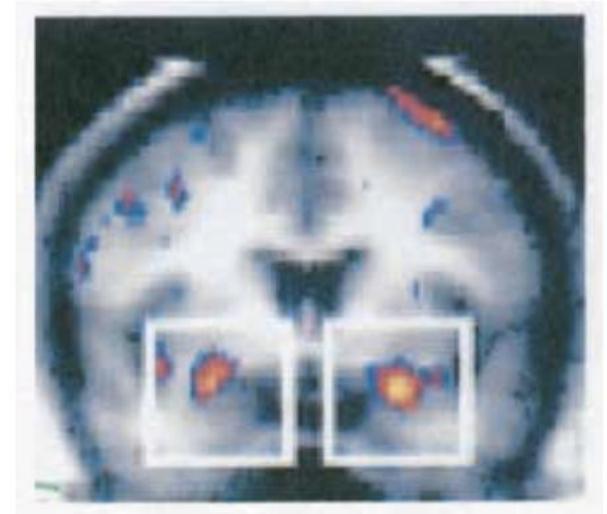
What happens if the intact amygdala is **electrically stimulated**?

Depending on the site, amygdala stimulation can lead to different effects, including a state of increased vigilance or attention. Stimulation of the **lateral portion of the amygdala** in cats can elicit a combination of fear and violent aggression. Electrical stimulation of the amygdala in humans has been reported to lead to anxiety and fear.

Neutral      Fear



(a)



Breiter et al., (1996). *Neuron*

# Amygdala and unconditioned (innate) fear

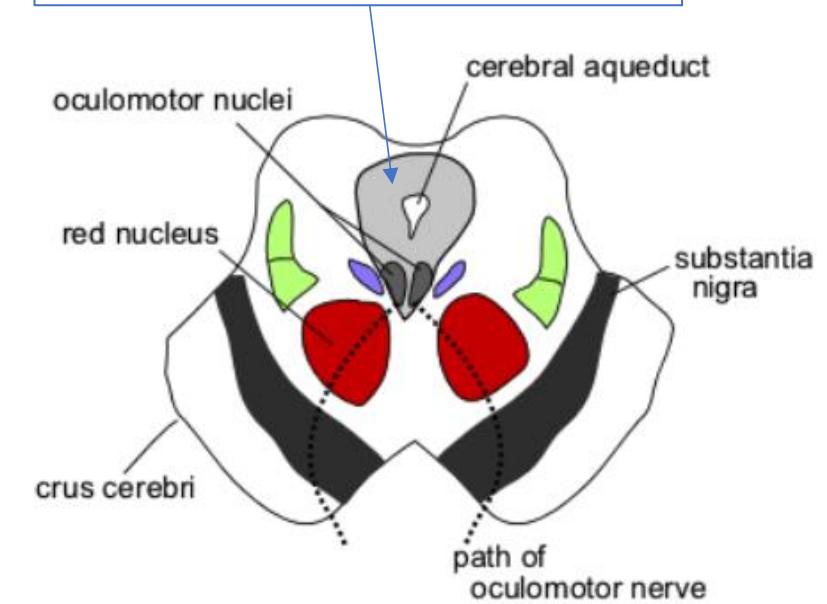
Many animals rely on **innate** (unconditioned) olfactory signals in the detection of threats, mates, food, and so forth. For example, rodents exhibit freezing and other defensive behaviors when fox urine is detected.

## Neural circuits underlying innate fear

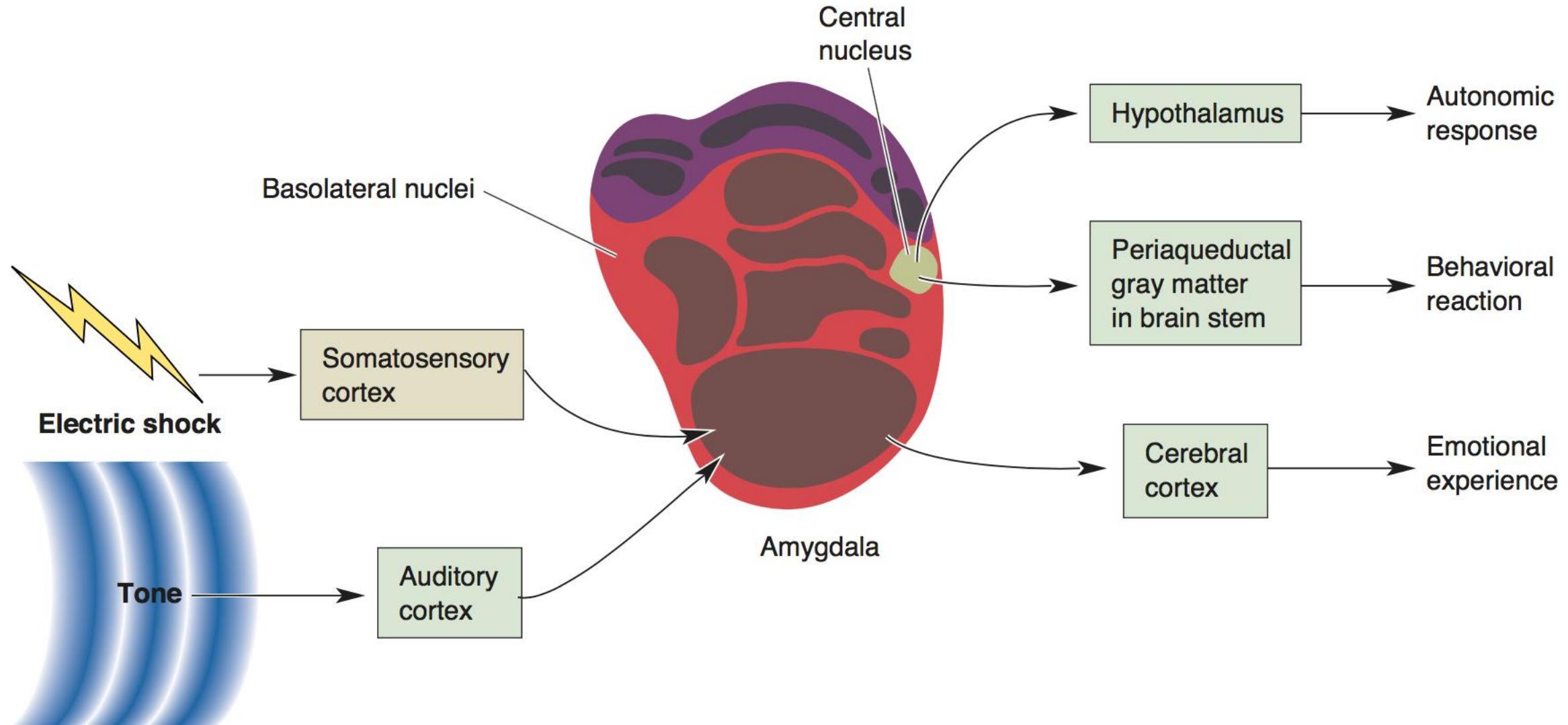
In mammals unconditioned threats involving predator or conspecific odors are transmitted from the **vermonasal component of the olfactory system** to **medial amygdala**. Outputs of the medial amygdala reach the **ventromedial hypothalamus**, which connects with the premammillary hypothalamic nucleus.

In contrast to learned fear, which depends on the **ventral periaqueductal gray (ventral PAG)** region, unconditioned fear responses depend on connections from the hypothalamus to the **dorsal PAG** region.

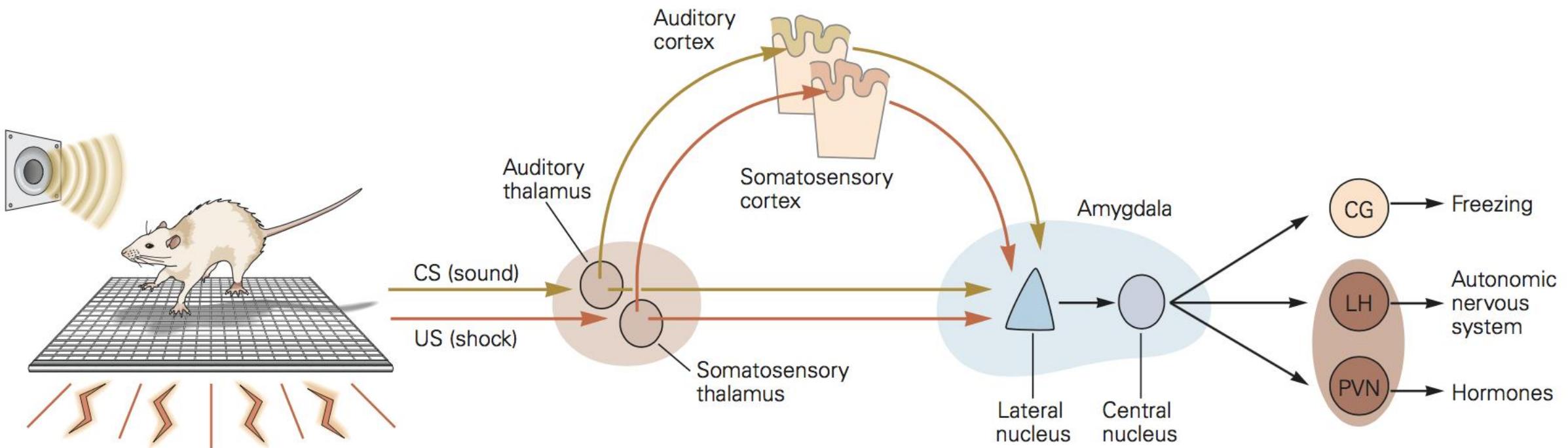
PAG is the gray area just peripheral to the cerebral aqueduct.



# Amygdala & learned fear (Fear-conditioning learning)



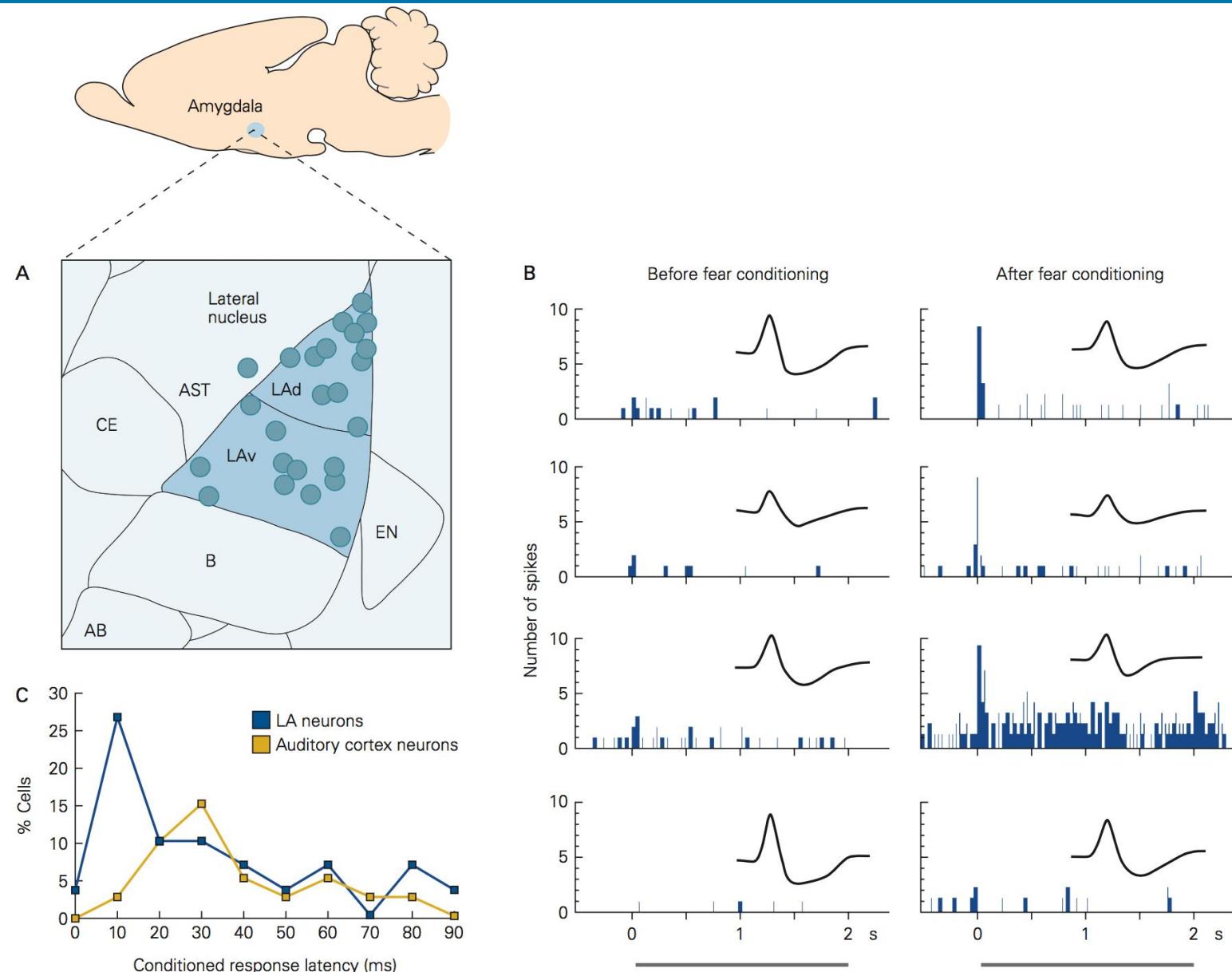
# Neural circuits engaged during fear conditioning



Medina JF, Repa CJ, Mauk MD, LeDoux JE. (2002). Parallels between cerebellum- and amygdala-dependent conditioning. *Nature Review Neuroscience*

# Amygdala & fear-conditioning learning

C. After conditioning, the neurons in the dorsal lateral nucleus are at **shorter latencies** than do neurons in the auditory cortex. They also fire at higher frequencies (not shown).

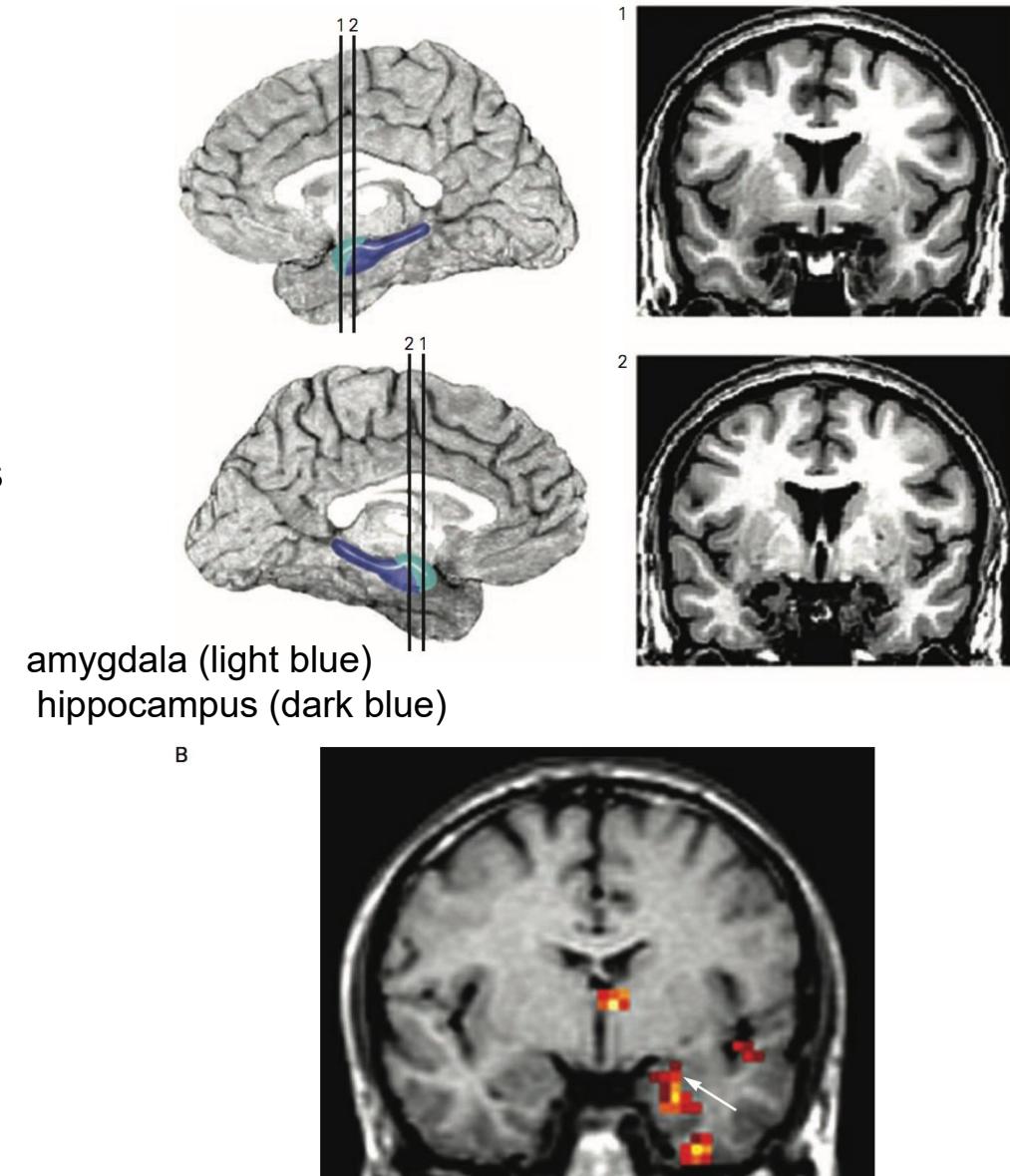


# Amygdala & fear-conditioning learning in humans

Certain forms of fear processing are **unique** to humans. For example, simply **telling** a human subject that the CS may be followed by a shock is enough to allow the CS to elicit fear responses. The CS elicits characteristic autonomic responses even though it was never associated with the delivery of the shock. Humans can also be conditioned by allowing them to **observe** someone else being conditioned—the observer learns to fear the CS even though the CS or US were never directly presented to the observing subject.

A. In such patients damage to the amygdala impairs fear responses and blocks fear conditioning.

B. *Conditioned fear stimuli* activate the human *amygdala* (arrow). Healthy volunteers underwent fear conditioning while their brains were scanned using functional MRI.



# Fear & anxiety in humans

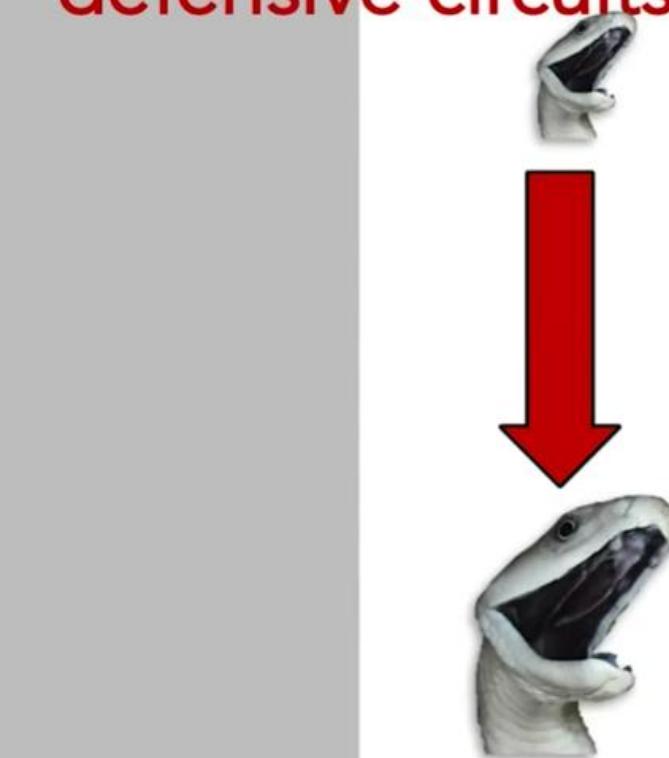
## Why study COVID-

- 31.1% of US adults report clinical anxiety
- Total annual mental health care visits to be between 100 and 150 million
- Health problems are increasing
- Negatively impacts mental health
  - 34% increase in depression during the outbreak.

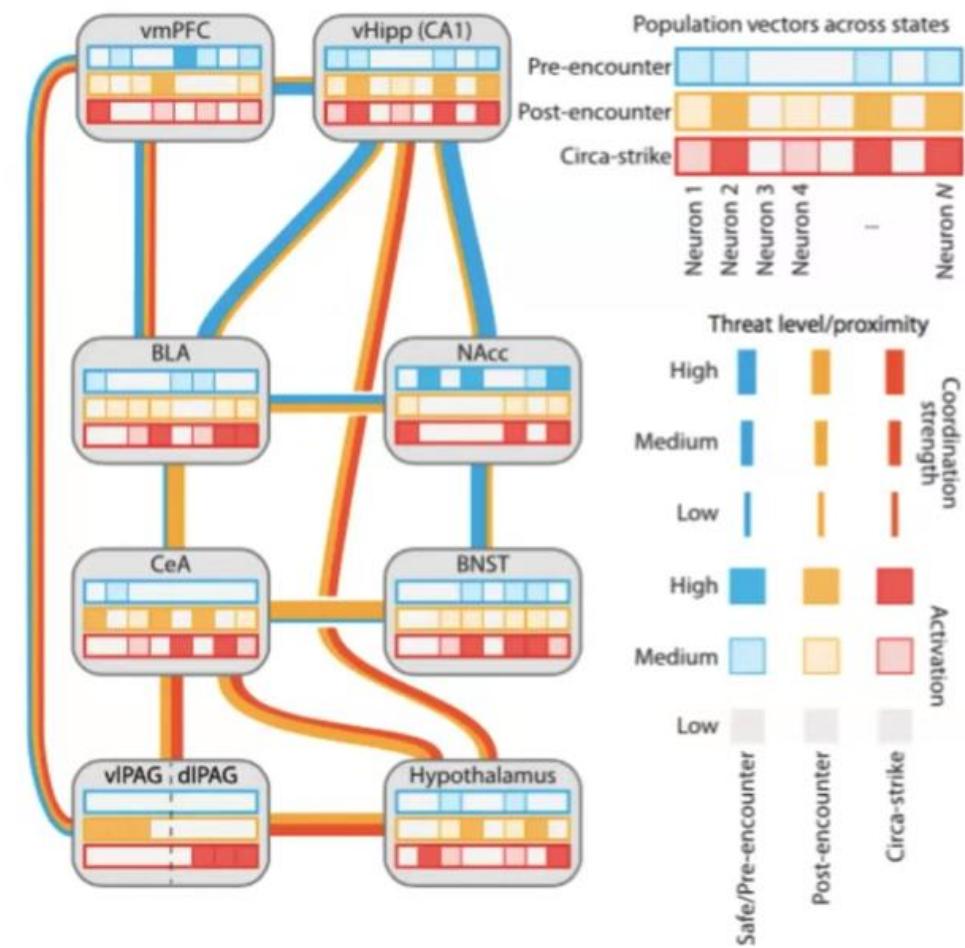
Van Bavel, J. J., et al.. (in press). COVID-19 and the psychology of pandemic response. *Nature Reviews Psychology*.

## Why study COVID-19?

## Population codes in defensive circuits



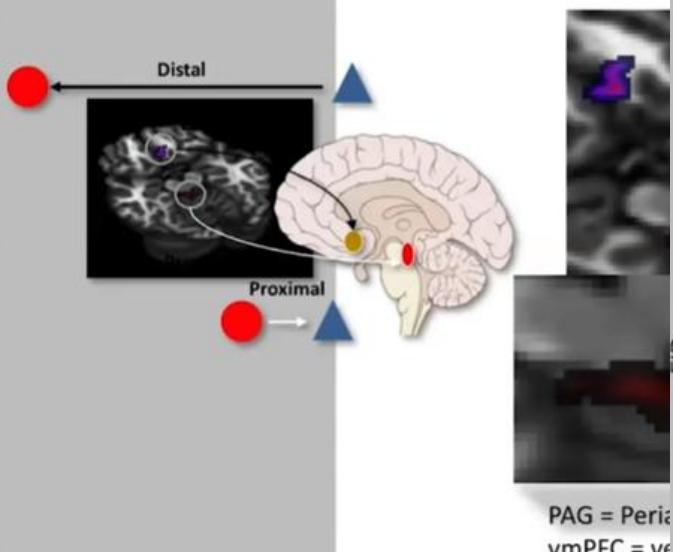
Mobbs et al., 2020, Trends in Cognitive Science



Prof. Dean Mobbs, The invisible predator: Fear and anxiety in the age of COVID-19. (2020)  
[https://www.youtube.com/watch?v=3\\_Wq-3-BI5U&t](https://www.youtube.com/watch?v=3_Wq-3-BI5U&t)

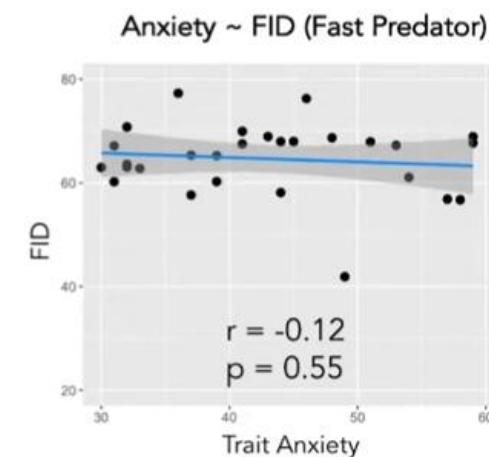
# Fear & anxiety in humans

## Active Escape Task



Mobbs et al., Science, 2007; Mobbs et al.

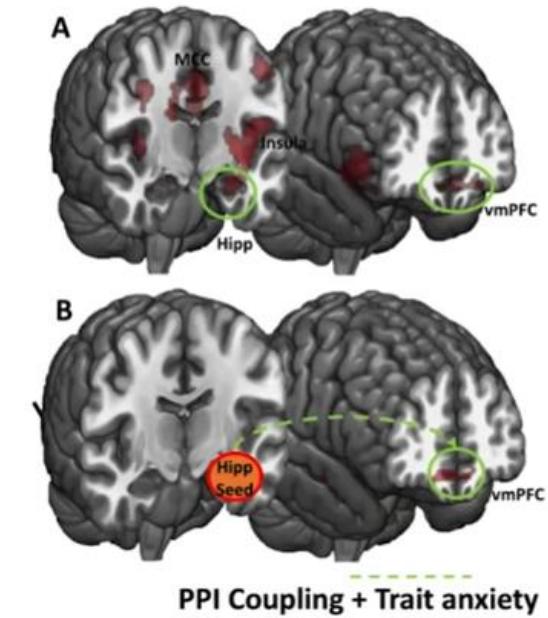
## Trait anxiety and flight distance



Fung, Qi et al, (2019) Nature: Human Behavior; Qi et al., PNAS.



Drs Bowen Fung/Song Qi



Caltech

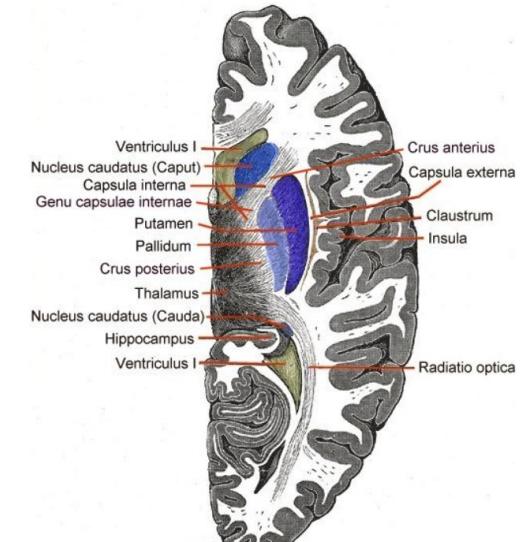
Prof. Dean Mobbs, The invisible predator: Fear and anxiety in the age of COVID-19. (2020)  
[https://www.youtube.com/watch?v=3\\_Wq-3-BI5U&t](https://www.youtube.com/watch?v=3_Wq-3-BI5U&t)

# Other Brain Areas contribute to Emotional Processing

- **Insular cortex**

Damage to the right somatosensory cortex (S-II, S-I, and insula) impairs social feelings such as **empathy**.

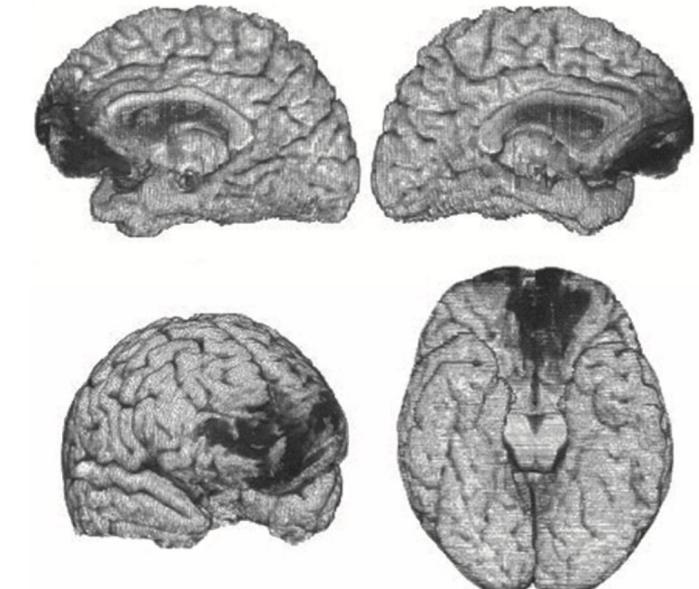
Damage to the human insular cortex, especially on the left, can **suspend addictive behaviors**, such as smoking.



- **Ventromedial prefrontal cortex (vmPFC)**

If vmPFC is damaged bilaterally, subjects have **impaired social emotions** making risky and disadvantageous decision

Immoral decisions



# Anger & aggression -- Hypothalamus, PAG, VTA

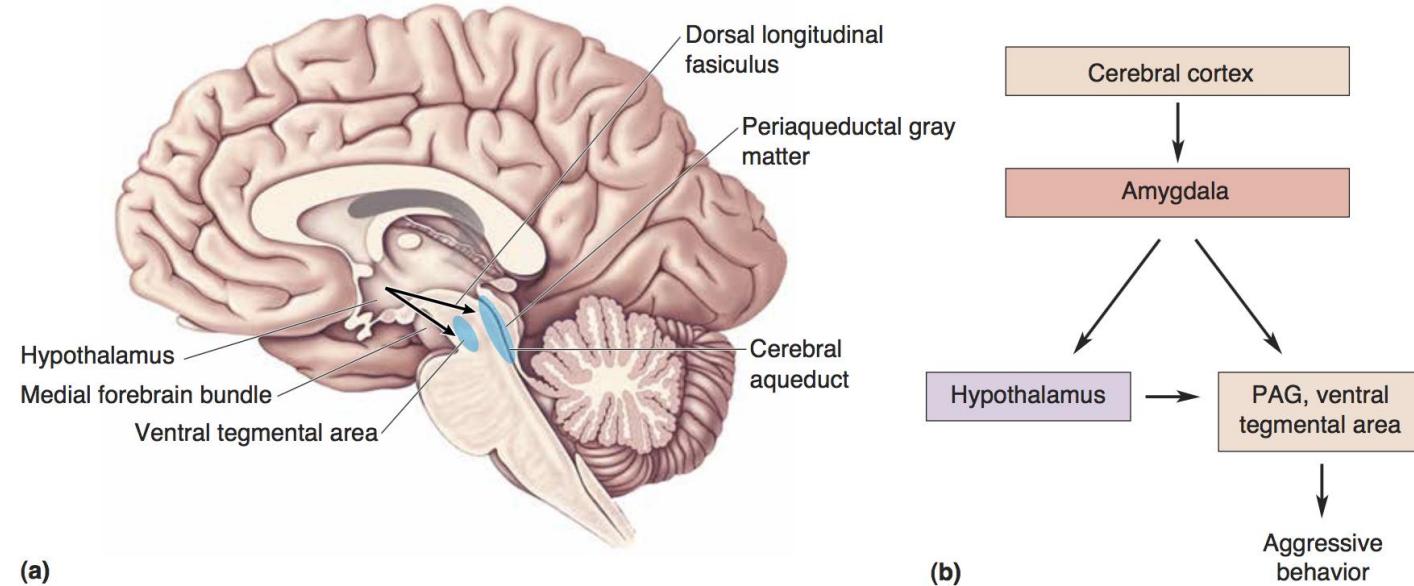
## Electrical stimulation to hypothalamus:

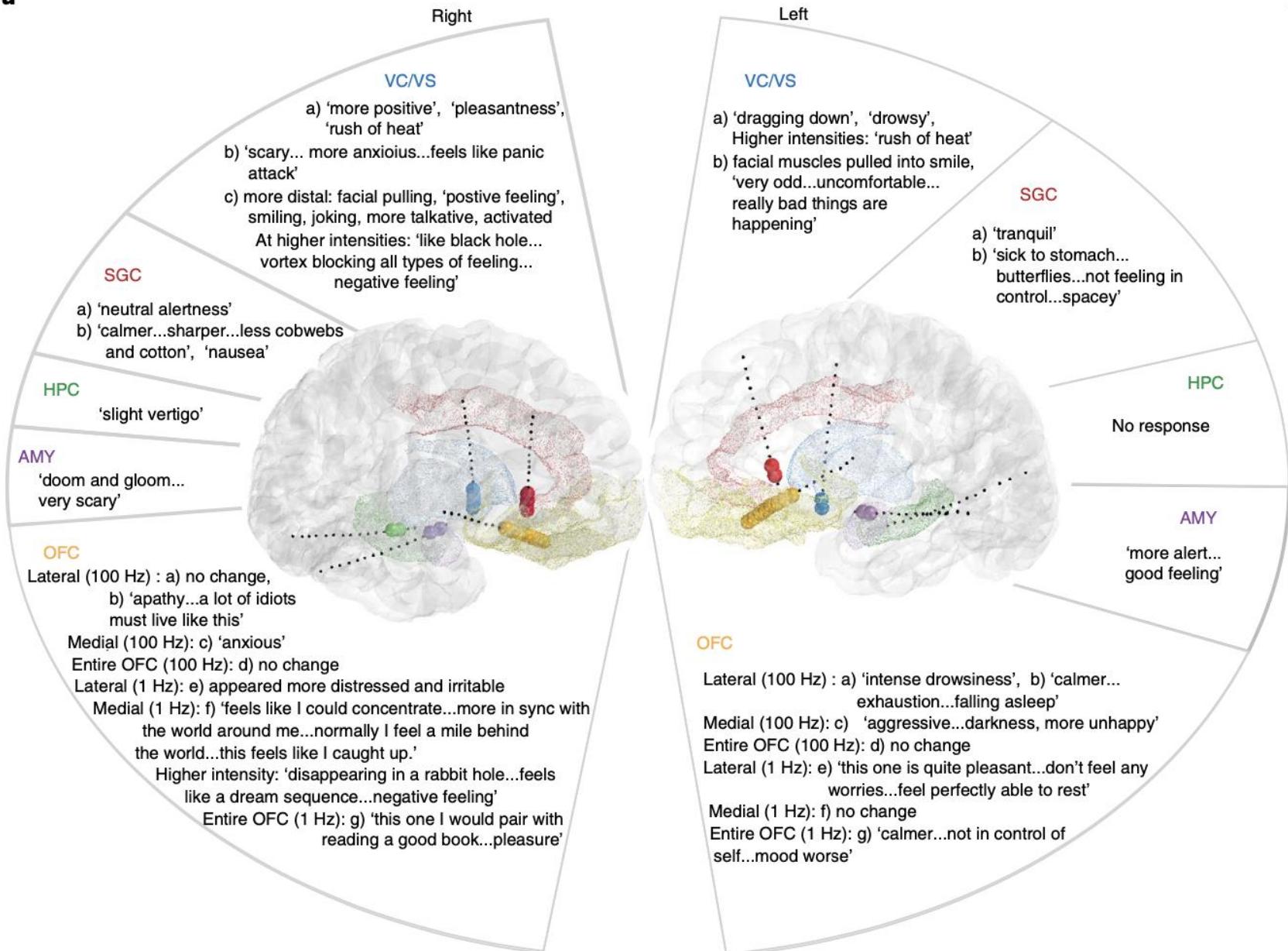
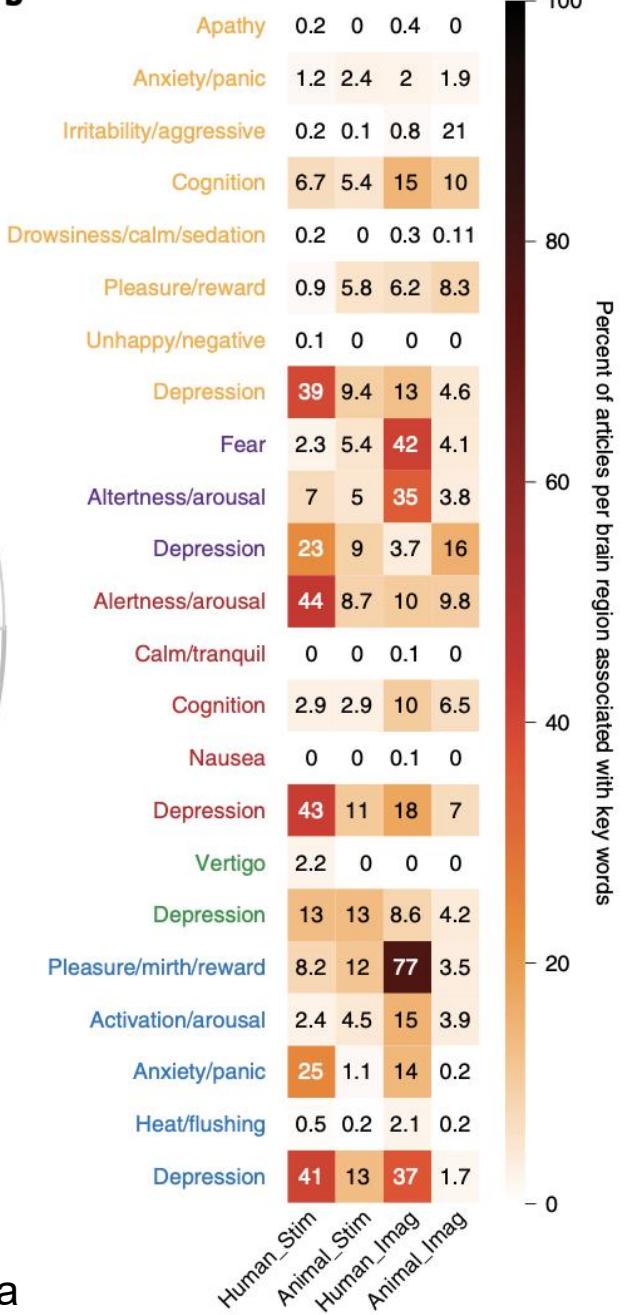
Depending on where the electrode is placed, stimulation may cause the animal to sniff, pant, eat, or express behaviors characteristic of fear or anger. These reactions illustrate the two primary functions of the hypothalamus: *homeostasis*; and the organization of coordinated *visceral and somatic motor responses*.

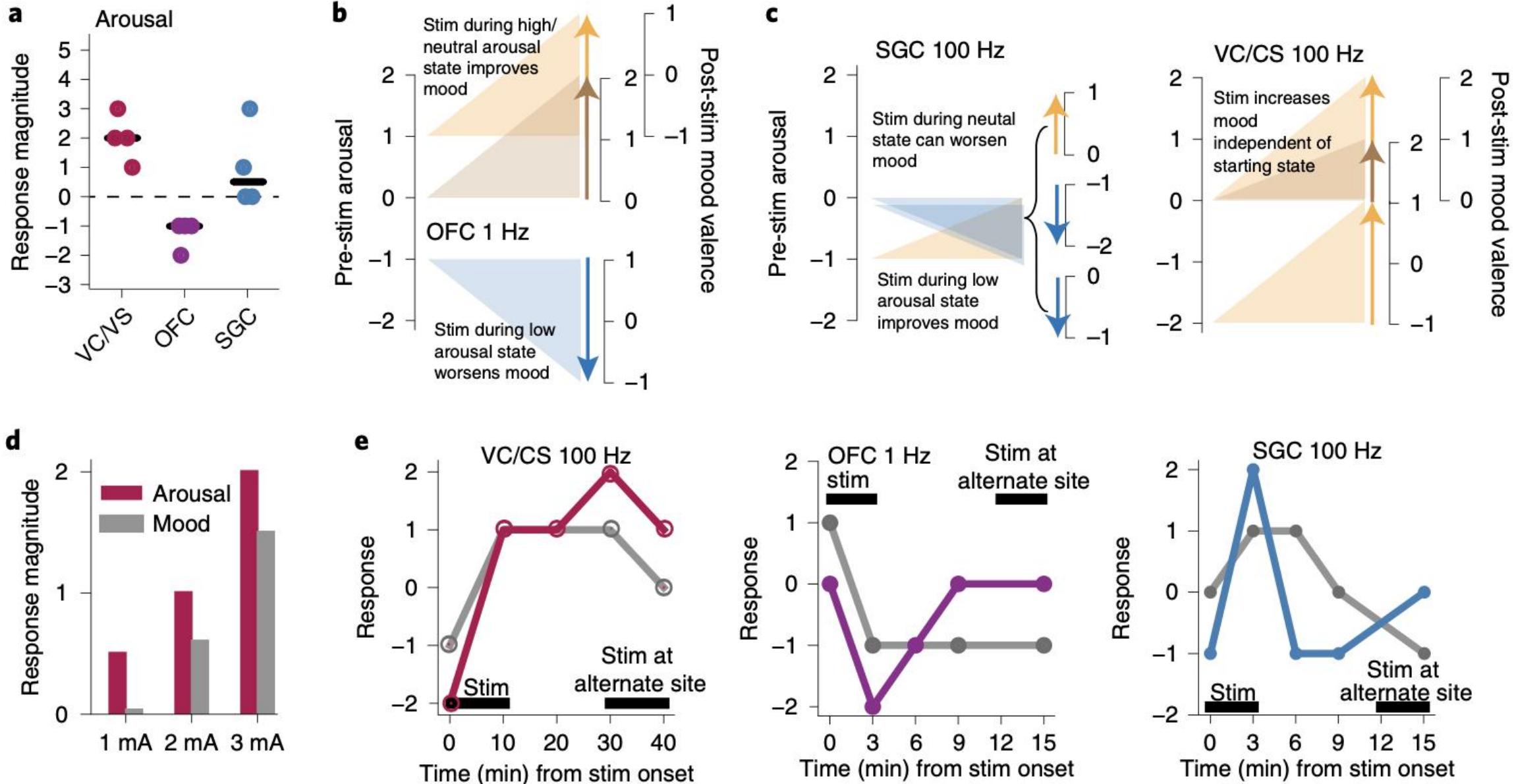
Stimulation of some parts of the hypothalamus also elicits behavior characteristic of **fear and anger**.

(a) The **hypothalamus** can influence aggressive behavior through projections to the periaqueductal gray matter (PAG), and ventral tegmental area (VTA).

(b) In this simplified scheme, the expression of anger and aggression is controlled by a neural pathway from the amygdala through the hypothalamus, PAG, VTA.

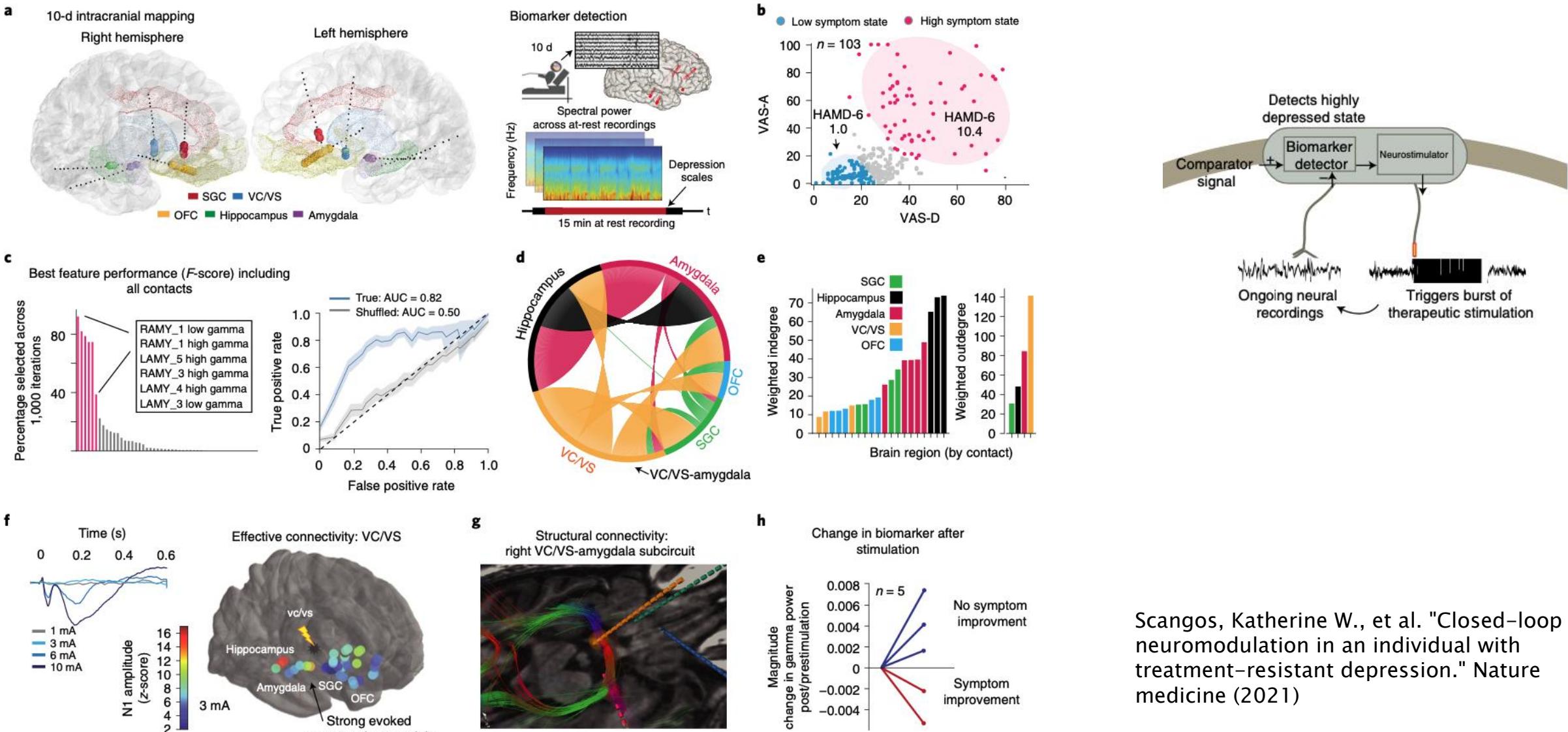


**a****b**



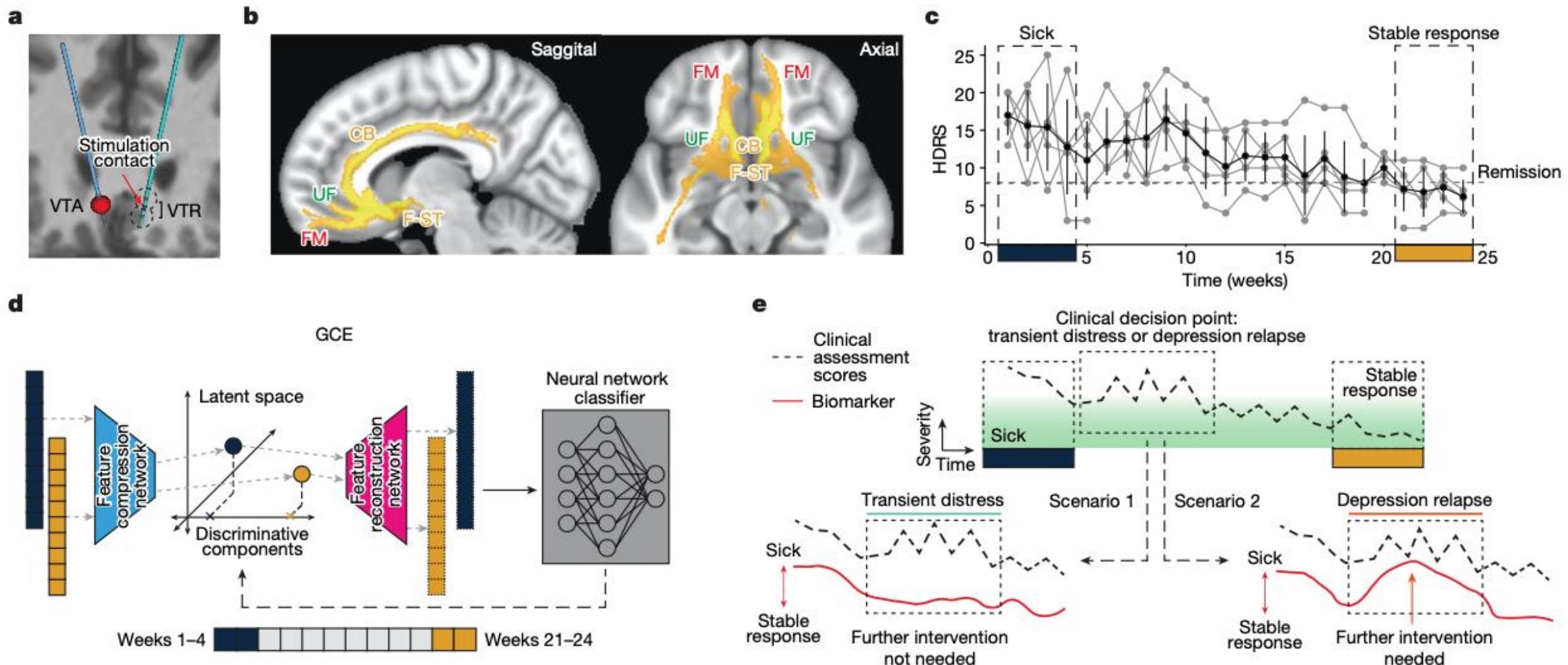
Scangos et al. (2021a) State-dependent responses to intracranial brain stimulation in a patient with depression, Nature Medicine,

# 闭环神经刺激治疗抑郁



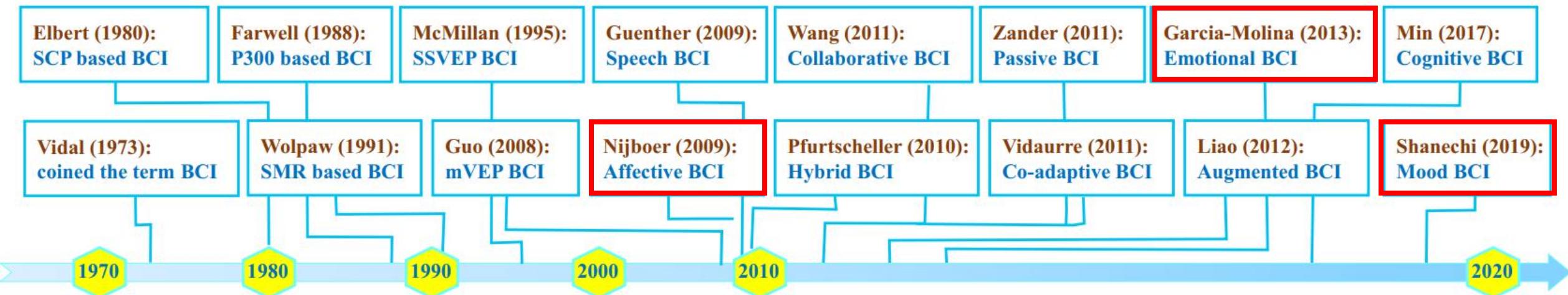
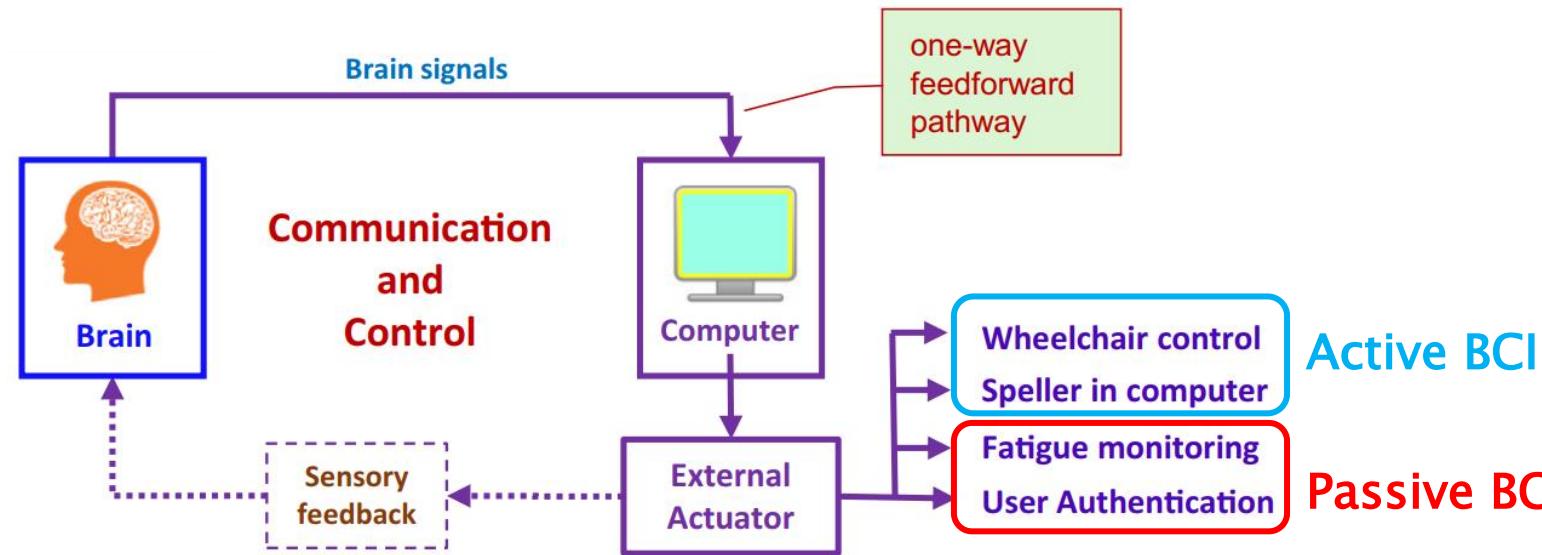
Scangos, Katherine W., et al. "Closed-loop neuromodulation in an individual with treatment-resistant depression." *Nature medicine* (2021)

# 神经活动预测DBS治疗抑郁的效果



# 情感脑机接口

情感脑机接口：一种对人的情绪进行识别和（或）调控的脑机接口（吕宝粮等，2021）。



# EEG datasets

- DEAP数据集 (Koelstra et al., 2012. 被引4197次)

<https://www.eecs.qmul.ac.uk/mmv/datasets/deap/>

情绪诱发素材 音乐视频

记录信号 32导脑电、眼电、肌电、皮电等

被试数量 32

情绪类型 高/低效价、高/低唤起

- SEED数据集 (Zheng & Lu, 2015. 被引1583次)

<https://bcmi.sjtu.edu.cn/home/seed/>

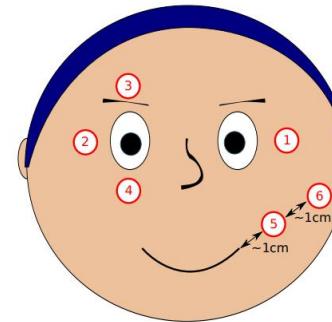
情绪诱发素材 电影片段

记录信号 62导脑电、眼动

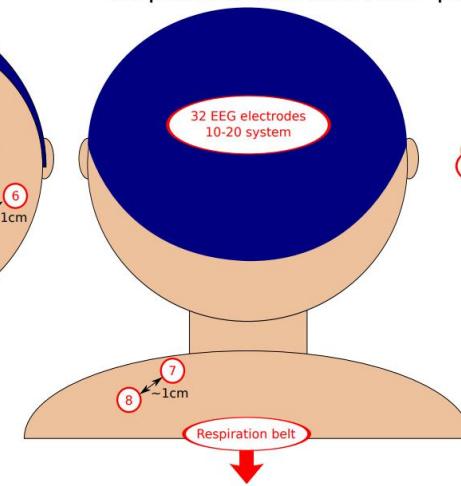
被试数量 15

情绪类型 积极、中性、消极

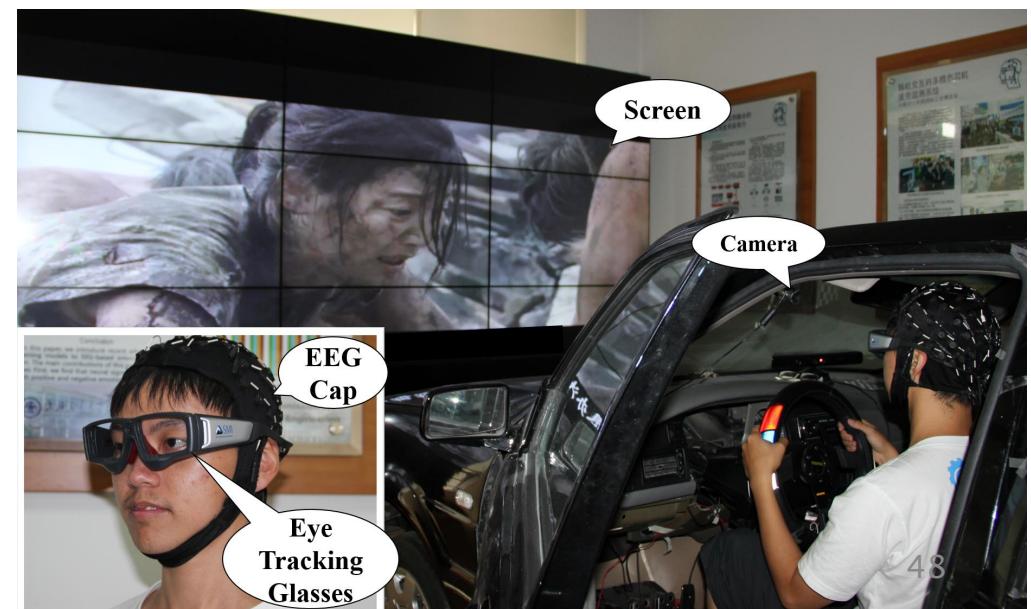
EXG sensors face



EXG sensors trapezius, respiration belt and EEG physiological sensors



Left hand physiological sensors



# EEG dataset from NCC lab



## • EmoEEG-MC dataset

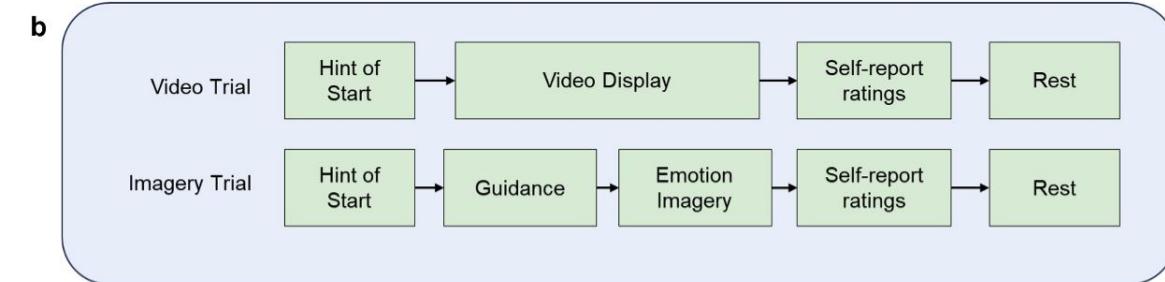
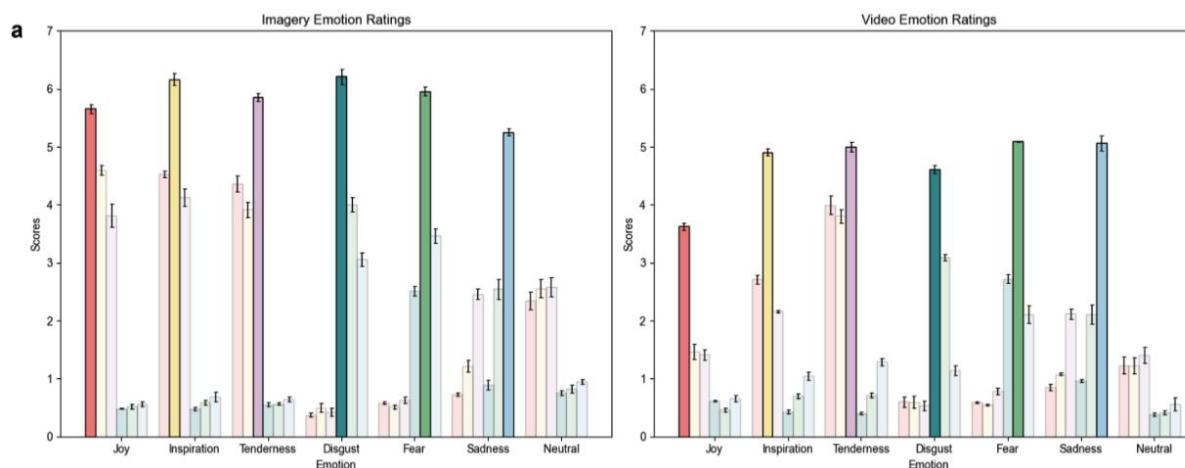
<https://github.com/ncclab-sustech/EmoEEG-MC>

情绪诱发素材 视频、想象

记录信号 64导脑电、眼电、肌电、皮电等

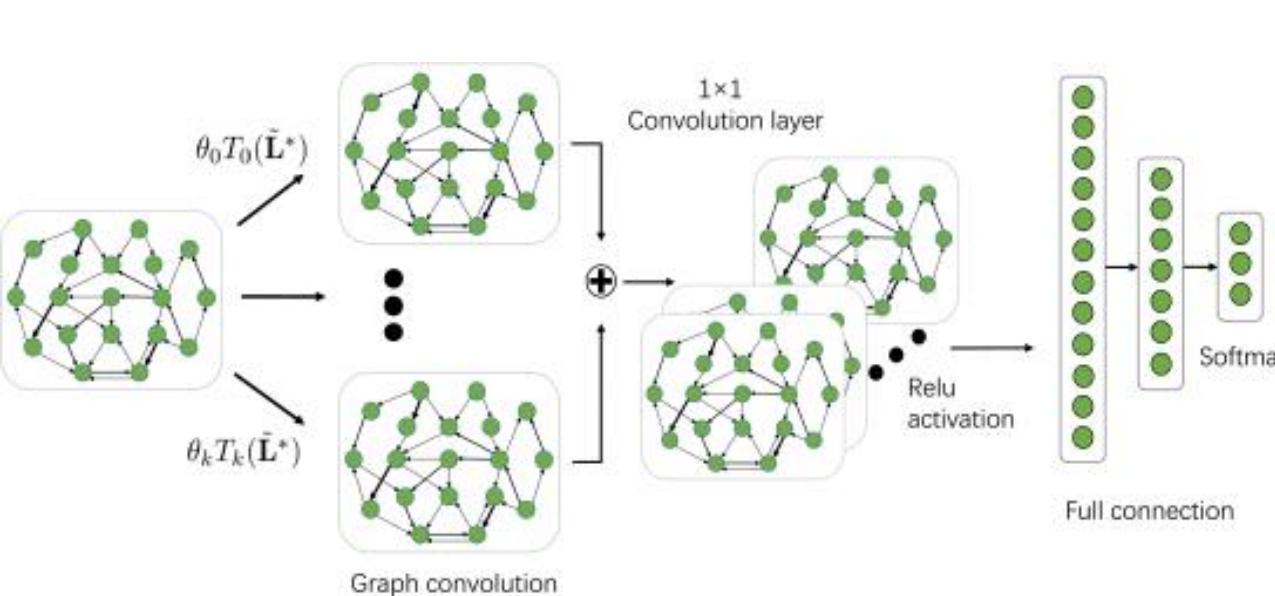
被试数量 63

情绪类型 6种情绪+neutral

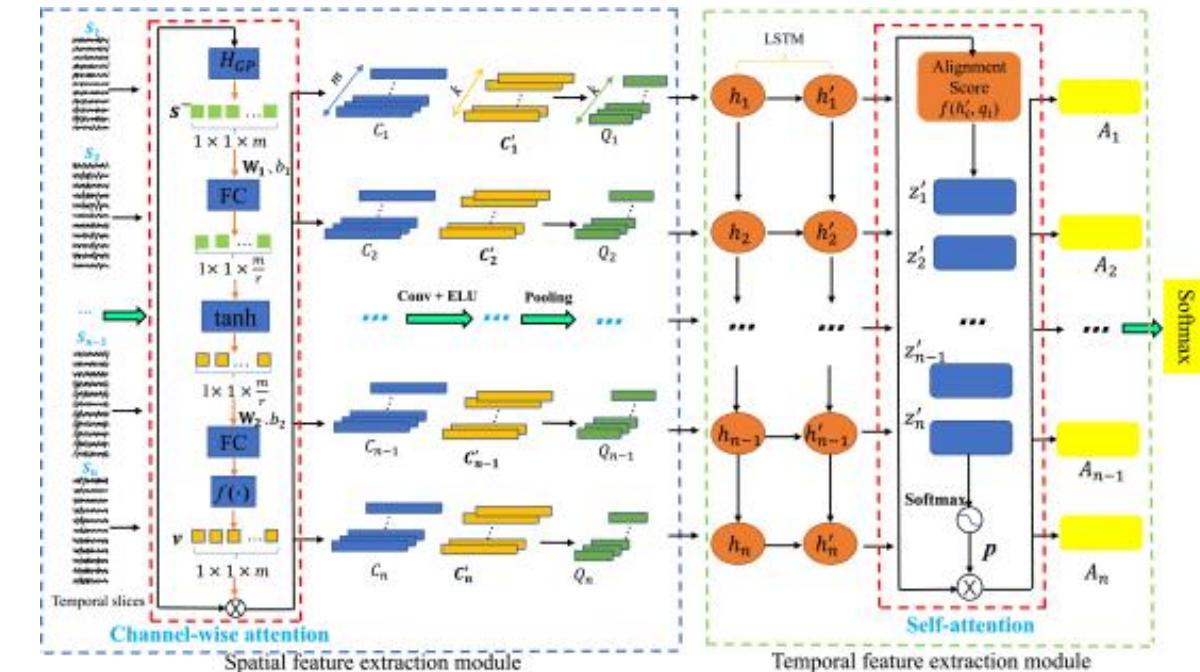


# 基于深度学习提升情绪识别的性能和泛化性

- 神经网络架构创新

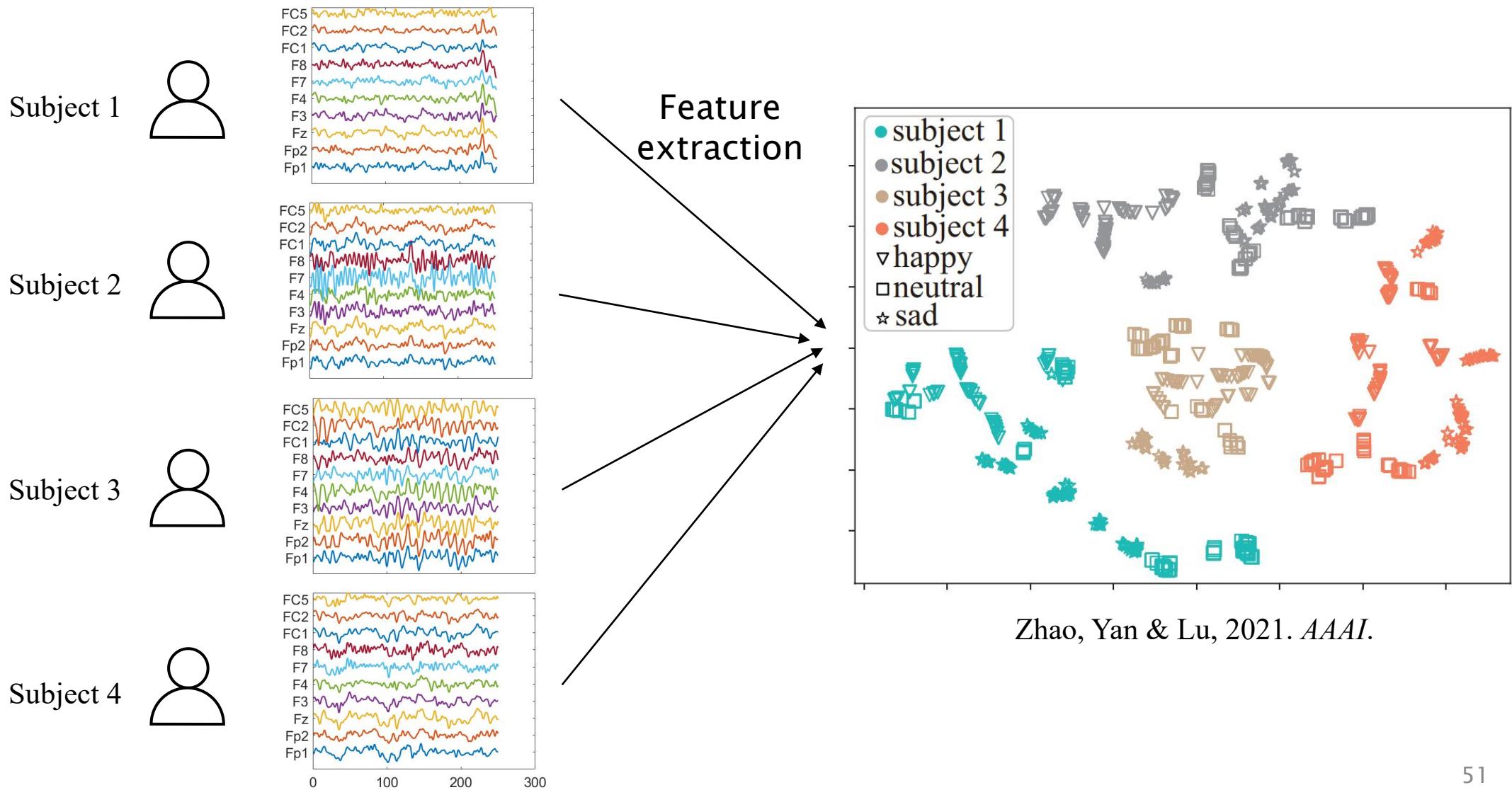


图神经网络 (Song et al., 2018. IEEE TAFFC)

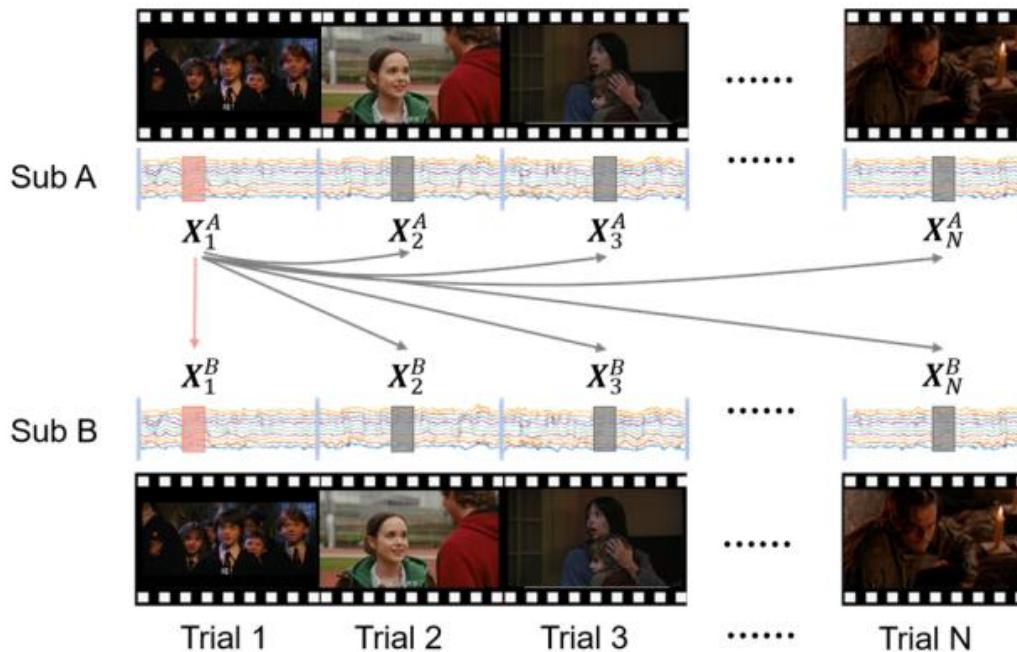


注意力机制 (Tao et al., 2020. IEEE TAFFC)

# 情绪相关特征在个体之间的不一致性

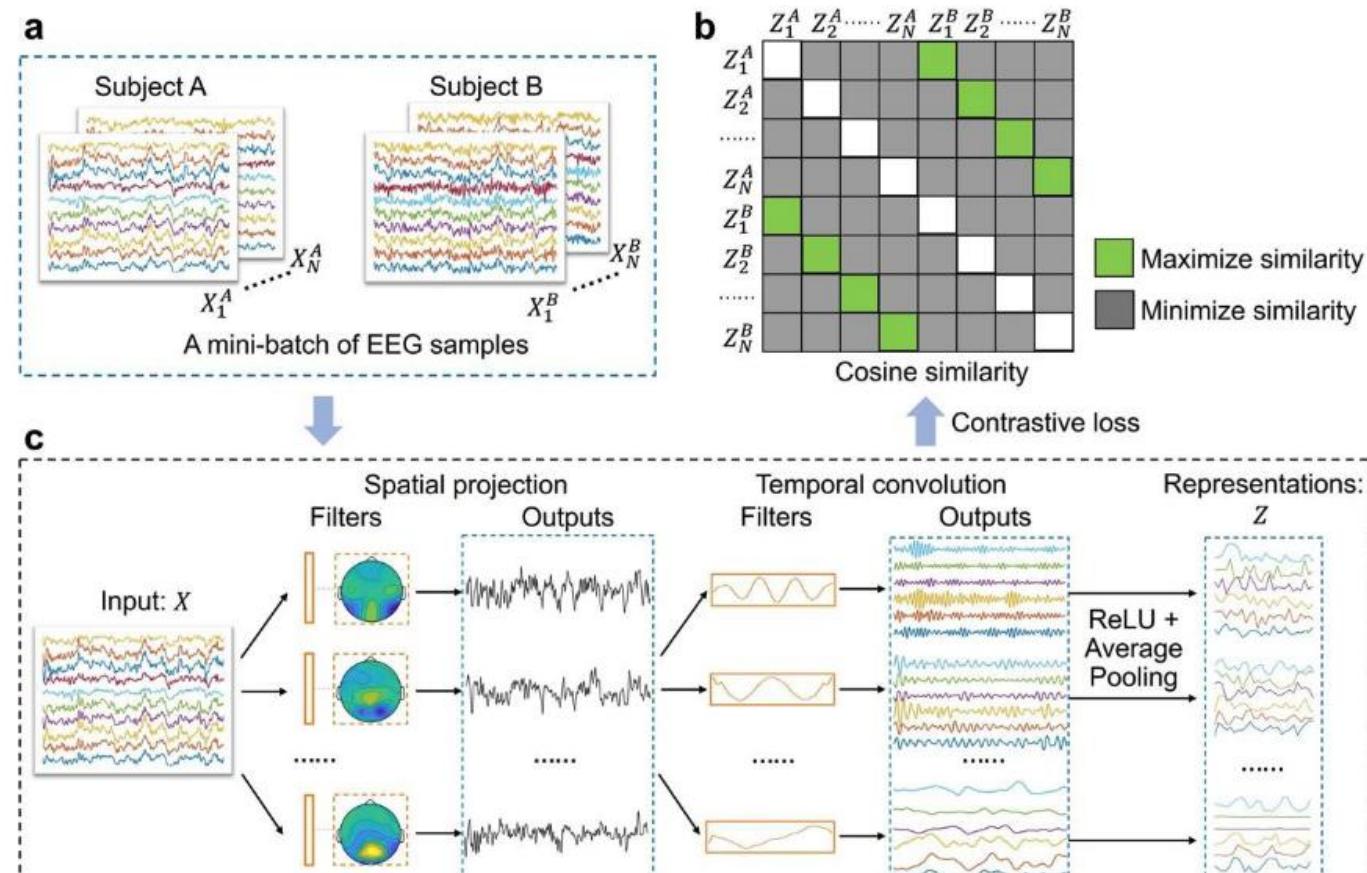


# 跨个体的情绪解码



对比学习方法

(Shen et al., 2022. *IEEE TAFFC*)



CLSSTER方法

Contrastive Learning of Shared SpatioTemporal EEG Representations

(Shen et al., 2024. *Neuroimage*)

# Emotional artificial intelligence

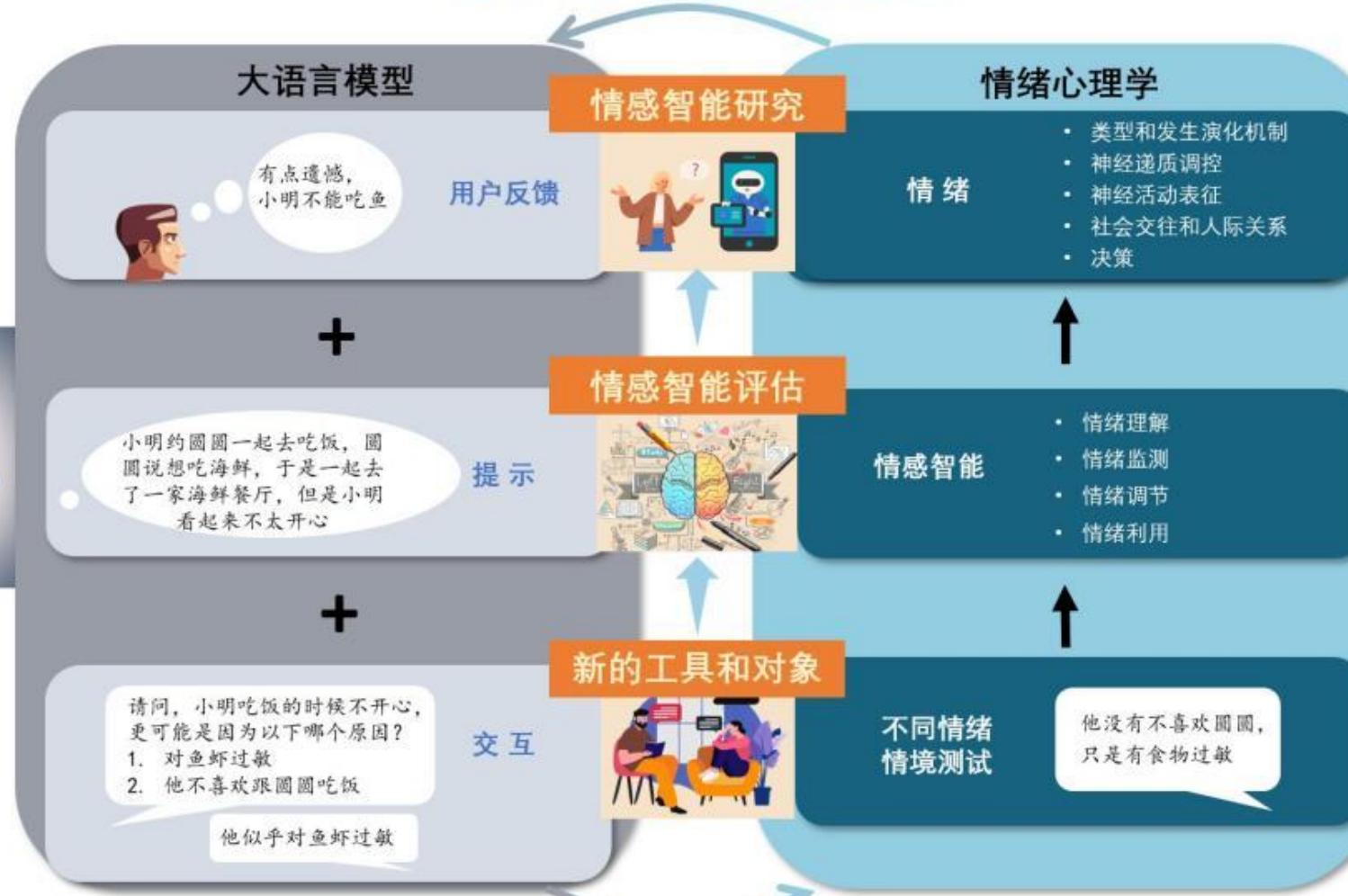
---

**Vote:**

1. Do you think **emotions** are **beneficial** to AI?
2. Do you agree that AI should be **rational**, rather than emotional?

# 大模型与情绪智能

## 心理学视角下的大模型情绪智能



与澳门大学伍海燕教授的合作

# Emotional artificial intelligence

## Dialoguernn: An attentive rnn for **emotion detection** in conversations

[N Majumder, S Poria, D Hazarika, R Mihalcea...](#) - Proceedings of the AAAI ..., 2019 - aaai.org

**Emotion detection** in conversations is a necessary step for a number of applications,

including op  
mining, und

## Robust real-time **emotion detection** system using cnn architecture

[S Jaiswal, GC Nandi](#) - Neural Computing and Applications, 2020 - Springer

... auto encoders are also used to build a model for classifvina 7 **emotions** in [16] ... A survev in various

## **emotion detect**

a brief review at

[☆ 99 Cited](#)

## Survey on **emotional body gesture recognition**

[F Noroozi, D Kaminska, C Corneanu...](#) - IEEE transactions on ..., 2018 - ieeexplore.ieee.org

Automatic **emotion recognition** has become a trending research topic in the past decade.

While works based on facial expressions or speech abound, recognising affect from body

to ...

## Automatic human **emotion recognition** in surveillance video

[J Arunnehr, MK Geetha](#) - Intelligent Techniques in Signal Processing for ..., 2017 - Springer

... [1] proposed human **emotion recognition** of four performed **emotional** states (angry ... are helpful for indexing and retrieving the multimedia information based on **emotion**-specific information

[13] ... an intelligent classifier that has the ability to discriminate four **emotions**, namely angry ...

[☆ 99 Cited by 21 Related articles All 3 versions](#)

# text / facial expression / voice中识别情绪



Emotion Detection in Python | **Emotion detection in 5 lines** | Machine Learning | Data Magic

2.3K views • 4 months ago

Data Magic

Hello Friends, Today we are going to do emotion detection for the face images with the help of DeepFace library which is ...

5:20

Deep Learning architecture  
Face Detection → Emotions  
Live Webcam Testing

Step By Step Implementation

32:24

This video contains python implementation of Realtime Face Emotion Recognition 1) Brainstorming (background of facial emotion ...)

# 情绪识别AI模型

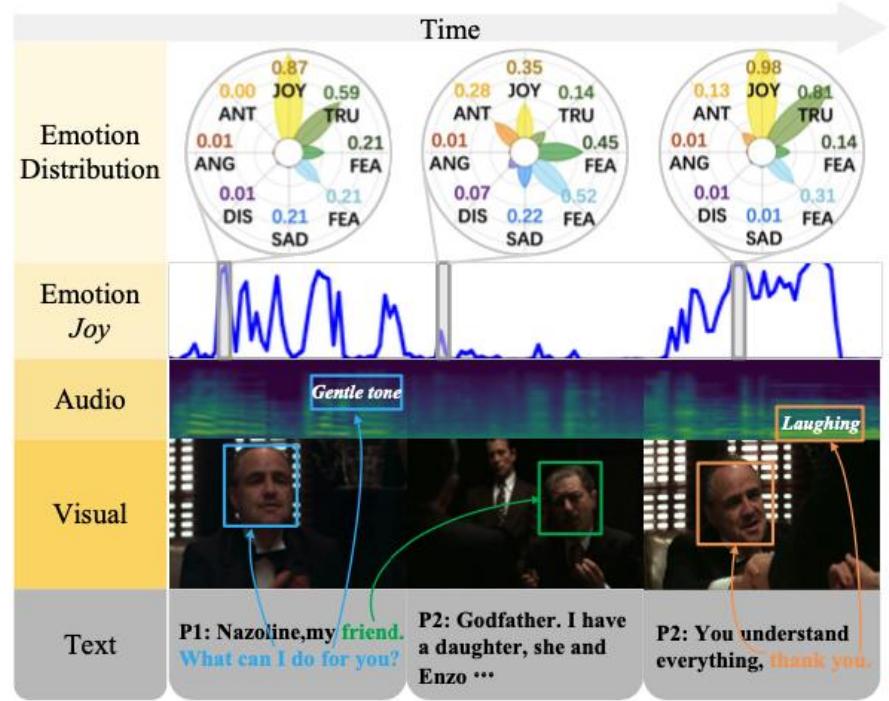


Figure 1. A segment from *The Godfather* (1972) and the corresponding emotion scores predicted by [84]. The colored texts indicate the emotional words given by looking up the affective lexicon [7] and the rectangles show the related video content towards emotional words.

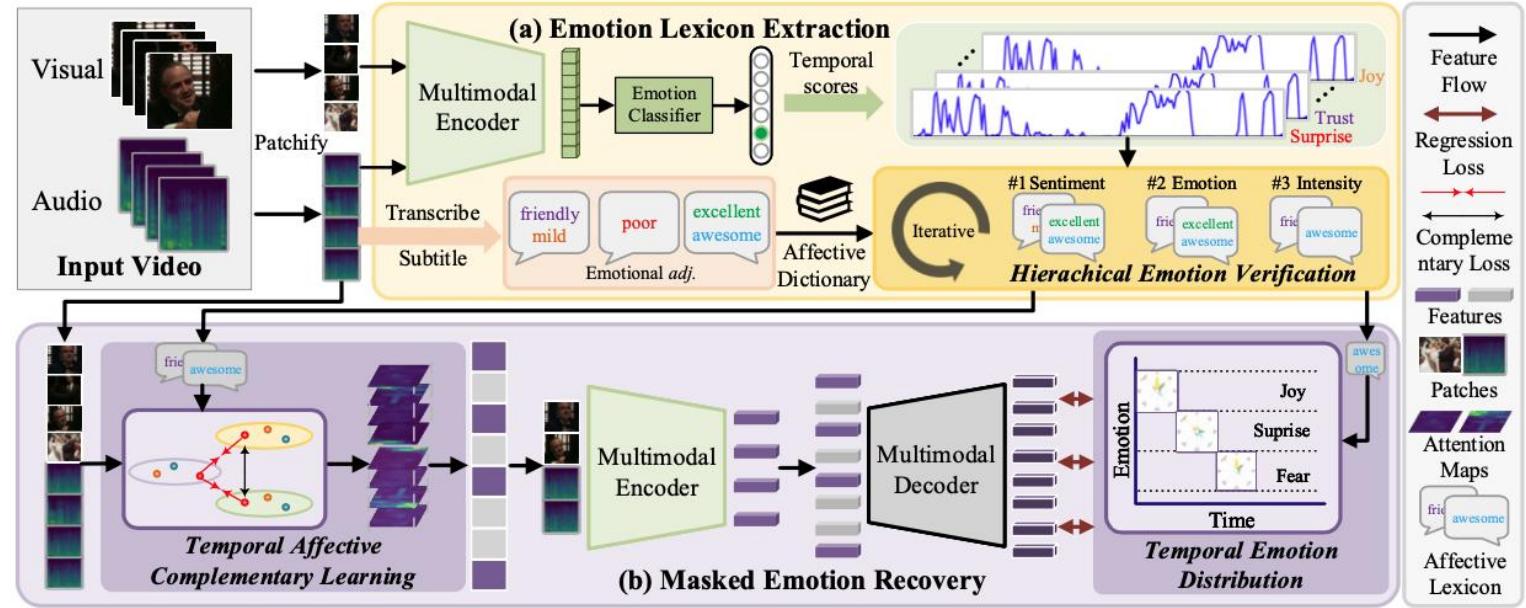


Figure 2. The pipeline of the proposed self-supervised method MART. Our work can be integrated into any existing SOTA VEA works by introducing our self-supervised learning method. Our method is decoupled into two modules for leveraging affective cues. In the emotion lexicon extraction module, we extract affective cues (*i.e.*, emotional words) and filter them via hierarchical lexicon verification. In the masked emotion recovery module, we first proposed to mask the part with strong emotion, which is measured by cross-attention between emotional words and video content. Then, we train the multimodal network to focus on affective segments of video and recover the corresponding temporal emotion distribution.

# EmoLLMs

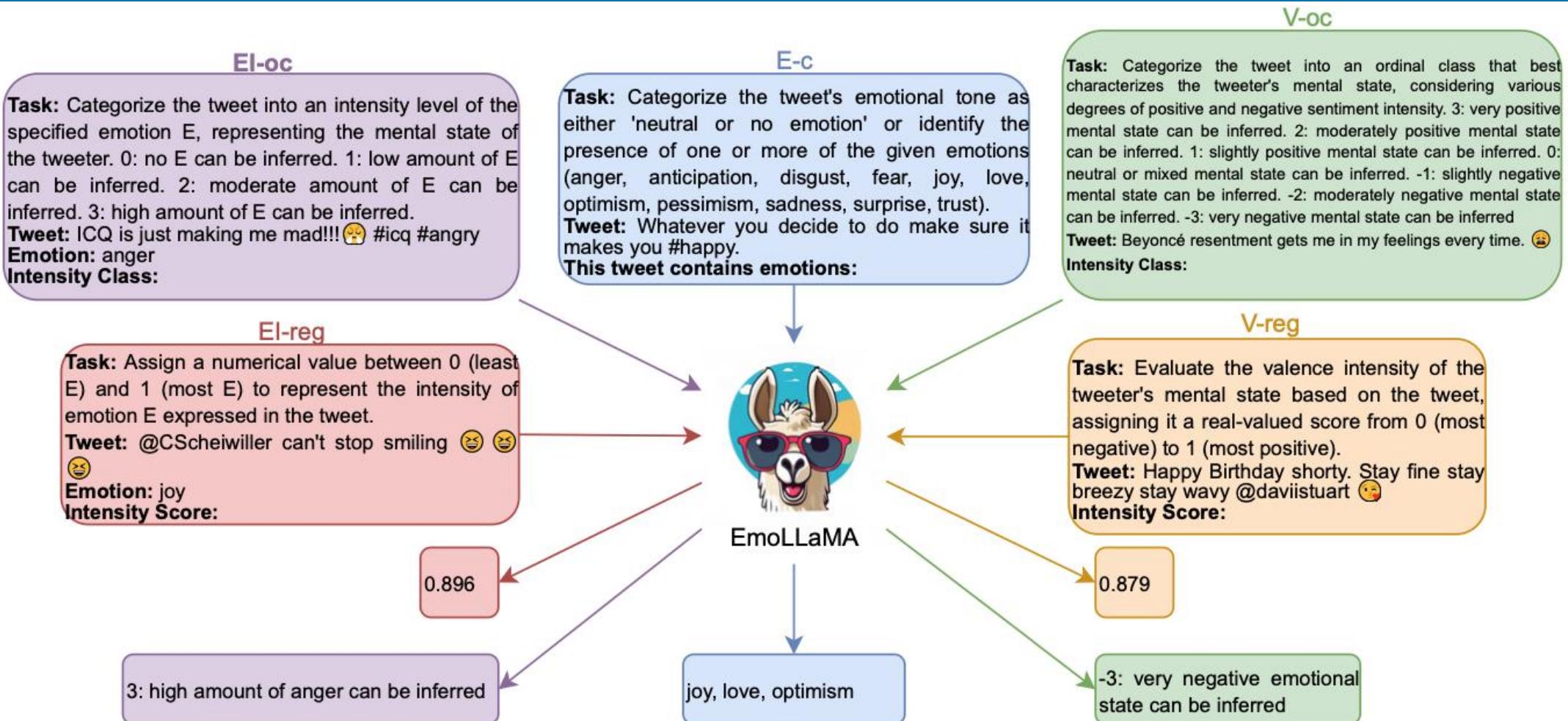


Figure 1: An overview of multi-task instruction tuning of EmoLLaMA for multiple affective analysis tasks.

# EmoLLMs

model	EI-reg					EI-oc					V-reg		V-oc		E-c	
	ave	anger	fear	joy	sadness	ave	anger	fear	joy	sadness	valence	valence	acc	mi-F1	ma-F1	
Leaderboard(1)	0.799	0.827	0.779	0.792	0.798	0.695	0.706	0.637	0.720	0.717	0.873	0.856	0.609	0.724	0.592	
<b>PLMs</b>																
BERT-base	0.785	0.800	0.781	0.783	0.742	0.683	0.698	0.656	0.712	0.665	0.840	0.805	0.567	0.718	0.568	
RoBERTa-base	0.717	0.670	0.736	0.769	0.694	0.664	-	-	-	-	0.845	0.772	0.563	0.721	0.536	
SentiBERT	0.722	0.724	0.740	0.731	0.691	0.665	-	-	-	-	0.835	0.763	0.535	0.700	0.522	
<b>Zero-shot/few-shot methods</b>																
Falcon	0.114	0.147	0.082	0.095	0.131	0.033	0.022	0.017	0.031	0.061	0.135	0.189	0.190	0.318	0.253	
Vicuna	0.281	0.307	0.257	0.260	0.299	0.214	0.238	0.193	0.186	0.241	0.298	0.579	0.220	0.359	0.253	
LLaMA-7B-chat	0.194	0.176	0.257	0.097	0.247	0.120	0.112	0.138	0.115	0.114	0.094	0.497	0.257	0.414	0.286	
LLaMA-13B-chat	0.488	0.524	0.506	0.398	0.526	0.194	0.262	0.178	0.119	0.216	0.312	0.568	0.274	0.424	0.302	
ChatGPT	0.599	0.637	0.573	0.569	0.618	0.455	0.500	0.428	0.363	0.529	0.637	0.748	0.382	0.546	0.429	
ChatGPT-FS	0.550	0.572	0.482	0.587	0.560	0.473	0.502	0.410	0.407	0.573	0.739	0.791	0.413	0.563	0.466	
GPT-4	0.656	0.699	0.575	<b>0.686</b>	0.667	<b>0.620</b>	<b>0.656</b>	<b>0.579</b>	<b>0.618</b>	<b>0.629</b>	0.811	0.788	0.444	0.572	0.497	
GPT-4-FS	<b>0.679</b>	<b>0.704</b>	<b>0.654</b>	0.679	<b>0.678</b>	0.562	0.623	0.523	0.515	0.585	<b>0.825</b>	<b>0.793</b>	<b>0.460</b>	<b>0.582</b>	<b>0.515</b>	
<b>Emotion-based instruction-tuning methods</b>																
EmoBART	0.795	0.798	0.803	0.795	0.782	0.725	0.705	0.742	0.723	0.729	0.851	0.835	0.528	0.686	0.548	
EmoT5	0.783	0.785	0.797	0.798	0.751	0.717	0.703	0.733	0.726	0.707	0.852	0.836	<b>0.559</b>	<b>0.712</b>	<b>0.568</b>	
EmoOPT	0.825	<b>0.827</b>	0.830	0.837	0.805	0.753	0.739	0.751	0.762	<b>0.759</b>	<b>0.887</b>	0.843	0.532	0.680	0.550	
EmoBLOOM	0.791	0.802	0.797	0.790	0.776	0.732	0.725	0.717	0.746	0.740	0.857	0.822	0.528	0.683	0.552	
EmoLLaMA-7B	0.822	0.819	0.821	0.837	0.809	0.743	0.738	0.722	0.768	0.745	0.879	0.843	0.545	0.695	0.563	
EmoLLaMA-chat-7B	0.824	0.825	0.830	0.832	0.810	0.751	0.748	0.754	0.764	0.739	0.876	0.827	0.534	0.693	0.540	
EmoLLaMA-chat-13B	<b>0.831</b>	<b>0.827</b>	<b>0.835</b>	<b>0.843</b>	<b>0.817</b>	<b>0.763</b>	<b>0.755</b>	<b>0.764</b>	<b>0.777</b>	0.755	0.886	<b>0.860</b>	0.537	0.696	0.545	

# Emotional artificial intelligence

Hume AI

## Ethical values

Our work is guided by The Hume Initiative

THE HUME INITIATIVE →

## AI ethics



### Beneficence

AI should be deployed only if its benefits substantially outweigh its costs.



### Empathy

AI privy to cues of our emotions should serve our emotional well-being.



### Scientific Legitimacy

Applications of AI should be supported by collaborative, rigorous, inclusive science.



### Emotional Primacy

AI should be prevented from treating human emotion as a means to an end.



### Inclusivity

The benefits of AI should be shared by people from diverse backgrounds.



### Transparency

People affected by AI should have enough data to make decisions about its use.



### Consent

AI should be deployed only with the informed consent of the people whom it affects.

# Quiz 3: Emotional AI (请作为人类进行独立思考)

一、请按照你心中觉得的重要性进行排序（从最重要到最不重要）

1. AI that can **understand** human emotions (through gesture, facial expression, voice, communications, and/or our brain signals)
2. AI that can make decisions that **considers and favors** human's emotion
3. AI that has its own emotions (**unconscious** physiological response)
4. AI that has its own emotions (**conscious** feeling)
5. AI that has its own emotions to **modulate** its own (AI's) decision making
6. AI that can **express** emotions to **modulate** human's emotion
7. AI that can **express** emotions in a way that human can **understand**

二、请按照你觉得的技术可行性进行排序（从最容易实现到最难实现），并解释原因

# GPT的回答：按重要性排序

Here's an ordered list with justifications based on their relevance to human-AI interaction, ethical concerns, and practical applications:

## 1. AI that can understand human emotions (through gesture, facial expression, voice, communications, and/or brain signals)

**Justification:** Emotionally intelligent AI systems, capable of interpreting human emotions accurately, are critical in healthcare, customer service, education, and personal assistance, where understanding emotions can significantly enhance user experience and outcomes. This form of AI has been shown to improve empathy in human-computer interactions, making it a top priority for real-world applications NINDS.

## 2. AI that can make decisions that consider and favor human emotions

**Justification:** Once an AI can understand human emotions, the next crucial step is to incorporate these insights into decision-making to enhance trust, comfort, and alignment with human well-being. Emotionally considerate decisions are particularly valuable in therapy bots, social robots, and any AI making decisions affecting vulnerable populations. The ability to make choices with empathy is a cornerstone of ethical AI design and could mitigate concerns of alienation in human-AI interactions.

## 3. AI that can express emotions in a way that humans can understand

**Justification:** Expression of emotions helps AI to better communicate intentions, enhance human trust, and provide reassurance, particularly in caregiving and education settings. Emotionally expressive AI may also help to reduce anxiety in users by offering comfort and motivation. This type of AI would need to display recognizable emotional cues, such as tone or gestures, that align with human expectations, making interactions smoother and more relatable.

## 4. AI that can express emotions to modulate human emotions

**Justification:** Beyond mere expression, AI designed to influence human emotions can be highly beneficial in therapeutic, educational, or customer service contexts, where the AI can encourage or calm users. However, influencing emotions ethically is complex, raising issues around manipulation, autonomy, and consent. Thus, this capability, while beneficial, requires rigorous ethical considerations.

## 5. AI that has its own emotions to modulate its own (AI's) decision making

**Justification:** AI that leverages its own "emotions" for internal modulation could be beneficial in adapting to different situations, particularly in dynamic environments where strict logic fails. Simulated emotions might improve decision-making processes and adaptability, much like how stress responses can guide humans to avoid risk. However, because these "emotions" aren't truly felt but are functional, their impact may be more limited in comparison to human-focused emotional intelligence.

## 6. AI that has its own emotions (unconscious physiological response)

**Justification:** Introducing unconscious "physiological" responses could give AI a closer resemblance to human reactions, making it feel more relatable and natural. Still, this is less important than actual interpretive or decision-making capabilities and could be a lower priority given its speculative utility.

## 7. AI that has its own emotions (conscious feeling)

**Justification:** AI possessing conscious feelings would raise fundamental ethical and philosophical concerns, including debates over AI rights and emotional welfare. Conscious emotions would imply subjective experiences, making AI potentially susceptible to harm or distress. This capability, while intriguing, is speculative and presents more ethical challenges than practical benefits, so it is placed as the least immediately necessary.

In sum, emotionally intelligent AI that can understand and make decisions based on human emotions ranks as the most impactful for current and near-future applications, while AI with its own emotions is less relevant for practical, ethical, and technical reasons.

# GPT的回答：按 技术可行性 排序

Here's an ordered list based on technical feasibility, from the easiest to most challenging to achieve, with explanations:

## 1. AI that can understand human emotions (through gesture, facial expression, voice, communications, and/or brain signals)

**Feasibility:** This is already widely feasible using natural language processing (NLP), facial recognition, and sentiment analysis. AI can interpret emotions through vocal tone, facial cues, and text with established algorithms (e.g., Convolutional Neural Networks for image processing in facial expression analysis and NLP models for textual sentiment analysis). Brain signal interpretation is more complex but advancing, especially with EEG data being used to interpret basic emotional states.

## 2. AI that can express emotions in a way that humans can understand

**Feasibility:** AI expression of emotions in a human-understandable form is technically straightforward. Text-based chatbots, synthesized speech with varied emotional tones, and simple facial expressions on robots or avatars can represent basic emotions like happiness, sadness, or anger. Examples include customer service chatbots that adjust tone based on user sentiment or robots designed to respond with simple gestures or sounds that humans interpret as emotional responses.

## 3. AI that can make decisions that consider and favor human emotions

**Feasibility:** This is moderately achievable with advancements in reinforcement learning and contextual decision-making. AI can already adjust responses based on detected emotions (e.g., de-escalating conversations when a user is frustrated). However, building a comprehensive decision-making model that reliably factors in nuanced emotions across diverse scenarios is complex and requires sophisticated contextual understanding.

## 4. AI that can express emotions to modulate human emotions

**Feasibility:** While AI can already express simple emotions, modulating human emotions intentionally and ethically requires a deep understanding of psychology and a personalized approach. Achieving this consistently involves modeling human responses in diverse situations, which requires substantial data and complex models for varying cultural, social, and individual contexts. It is achievable but demands a high degree of sensitivity and adaptability.

## 5. AI that has its own emotions to modulate its own (AI's) decision making

**Feasibility:** Developing an AI with simulated emotional states influencing its decision-making is possible with limited functionality, using internal states like "stress" or "confidence" that impact actions under certain conditions. Some autonomous systems use similar concepts to handle dynamic environments. However, engineering a system that continuously and adaptively modulates decisions based on these internal states to simulate emotion-driven behavior is challenging and requires advanced AI that closely resembles biological systems.

## 6. AI that has its own emotions (unconscious physiological response)

**Feasibility:** Creating an AI that mirrors unconscious, physiological responses seen in human emotion (e.g., increased heart rate) would necessitate new architectures or hybrid models that can simulate these reactions in a non-biological system. While it's possible to simulate these states through proxy signals, such as varying resource allocation or computational load, replicating the complex feedback systems of human physiology remains a significant challenge.

## 7. AI that has its own emotions (conscious feeling)

**Feasibility:** This is currently the least feasible, as it would involve not only replicating human-like conscious experiences but also defining and understanding "consciousness" itself—an area where neuroscience and AI remain in early stages. This would require breakthroughs in understanding subjective experiences and creating AI architectures with self-awareness, which is far beyond current capabilities.

In summary, interpreting and responding to human emotions is within current AI technology's reach, while creating systems with self-driven, conscious emotional states is far less achievable given our limited understanding of consciousness and emotion simulation.

# Summary of Lecture 13 – Emotion in Brain & AI

---

- **What is emotion?**
- **Emotion theory**
  - Basic emotion theory --> Dimensional emotion theory
  - Theory-driven --> Data-driven
- **Limbic system: amygdala & fear**
  - Innate fear
  - Learned fear
- **Other brain areas**
  - Insular & ventromedial prefrontal cortex (vmPFC)
  - Hypothalamus & brainstem (e.g. PAG)
- **Emotional EEG datasets**
- **Emotional AI**

# Reading materials

---

- *Neuroscience: Exploring the brain* (3<sup>rd</sup> ed), **Chapter 18 – Brain Control of Movement**, pp615-643
- *Principles of Neural Science* (5<sup>th</sup> ed), **Chapter 48 – Emotions and Feelings**, pp1079-1093
- Empathy Meets AI: A Fresh, More Humane Approach | Alan Cowen (Hume AI) | TransformX 2022

<https://www.youtube.com/watch?v=QZQwKa1sBIY>