



**Universidad Nacional Autónoma de México**  
Maestría en Ciencias (Neurobiología)  
Instituto de Neurobiología  
Departamento de Neurobiología Conductual y Cognitiva

## Estudio de los correlatos neurales de la percepción emocional por análisis de patrones en multitud de voxels

Que para optar por el grado de:

Maestro en Ciencias  
**Isaac David Reyes González**

Tutor principal:

**Dr. Fernando Alejandro Barrios Álvarez**  
Instituto de Neurobiología, UNAM

Comité tutor:

**Dr. Eduardo Adrián Garza Villarreal**  
Instituto de Neurobiología, UNAM  
**Dr. Luis Concha Loyola**  
Instituto de Neurobiología, UNAM

México, junio de 2020

# Abstract

Metanalyses in affective neuroscience have failed to agree on the specific neural substrates of distinct emotional experiences. Task-related fMRI has long relied on multiple-regression analysis of single-voxel signals (notably GLM), which assumes that activity at some region will correlate with experimental variables, but this needs not be the case: cognitive states might emerge from the *joint* activity probability distribution of many elements. This has led many to believe that emotion decoding could benefit from the more sensitive machine-learning methods of multi-voxel pattern analysis (MVPA). However, so far these have only been used to either explore one region at a time, or with no intention to derive group-level activation imaging when used brain-widely. This work explores the extent to which MVPA can overcome these limits as an actual substitute for GLM, and applies it to the problem of narrowing down the neural substrates of emotion *perception*.

Two similar experiments were performed, each one using a separate validated stimuli dataset on a different control group of people, who underwent their respective emotion perception task on a 3 T MR scanner. Subjects were presented with neutral, happy, sad and angry (second experiment only) faces in equal amounts at random. For the first experiment, short blocks of interest were interspersed with a variable amount of neutral control blocks, so as to avoid both subject habituation and test-partition contamination by temporal correlation. The second experiment addressed complexity and confounding concerns in the former, and amassed 800 % more training data.

After image preprocessing, linear SVM classifiers (one per subject) were trained for all non-trivial block combinations. Predictive power was estimated using cross-validation, then compared to the null hypothesis distribution estimated using the Monte-Carlo method (5000 training-partition label permutations). Instead of assuming an optimal labeling delay after stimulus presentation, the timeseries of classification accuracy were measured for all possible delays given fMRI's temporal resolution; and group-level hypothesis testing was conducted considering the single delay that maximized average accuracy for a particular class contrast (first and second experiment), and also the best delay per subject (first experiment only, corrected for multiple comparisons).

Average classification timeseries' morphology agrees with theoretical expectations about the hemodynamic response, although little to no evidence of successful decoding was found for the first database; even after some attempts at dimensionality reduction (top 1 % of gray-matter voxels according to their F-test score), which pushed 3-way classification performance to 35 % accuracy (delays: 2.2 s and 5.2 s). When allowing different delays

per subject, accuracy surpassed the 50 % line to great statistical effect ( $D_{Cohen} = 3,7$ ), but was likely to be a coincidence ( $p_{Bonferroni} = ,22$ ).

Depending on emotion combination, average classification timeseries for the second database saw most maxima occur 4 s after stimulus onset, and never immediately or too late. Performance varied from random guessing to statistical significance; yet this variability followed a pattern: only all combinations including *happiness* showed evidence of successful decoding. E.g. happy vs angry ( $p = ,04$ ,  $D_{Cohen} = 2,2$ ), happy vs sad ( $p = ,07$ ,  $D_{Cohen} = 1,9$ ), happy vs sad vs neutral ( $p = ,09$ ,  $D_{Cohen} = 1,7$ ). No dimensionality reduction was performed other than gray-matter masking. Moreover, statistical mapping ( $z\_value > 2,3$ ) of normalized SVM weights for successful contrasts consistently threw structures implicated in facial and emotional processing: anterior cingulate, orbitofrontal, insular and fusiform cortices; amygdalas and ventral striatum. Meanwhile, classical analysis with GLM was unable to find any significant activation cluster.

These results indicate that standard fMRI techniques can record differential information pertaining to brain activity associated with certain emotions (notably happiness), so a multivariate algorithm like SVM could learn to decode such slippery states above chance levels, whereas the univariate algorithm couldn't; statistical power being equal. The anatomical consistency of discriminating features across subjects and contrasts despite high-dimensionality and a small subject sample, as well as their correspondence with the wider literature, suggests MVPA may be a viable tool for functional neuroanatomical mapping and not just prediction of psychological states.

# Índice general

<b>Abstract</b>	<b>I</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Antecedentes</b>	<b>8</b>
2.1. Bases neurales de las emociones . . . . .	8
2.2. Resonancia magnética y el efecto BOLD . . . . .	13
2.3. Análisis estadístico de patrones de actividad cerebral . . . . .	20
2.3.1. Análisis clásico . . . . .	20
2.3.2. Neuroimagen funcional como un sistema dinámico . . . . .	25
2.3.3. Mapas de activación: análisis clásico vs MVPA . . . . .	26
2.3.4. MVPA en neuroimagen funcional afectiva . . . . .	29
<b>3. Justificación</b>	<b>33</b>
<b>4. Hipótesis</b>	<b>37</b>
4.1. Preguntas de investigación . . . . .	37
4.2. Hipótesis . . . . .	37
4.3. Predicciones . . . . .	37
<b>5. Objetivos</b>	<b>38</b>
5.1. General . . . . .	38
5.2. Específicos . . . . .	38
<b>6. Muestra, Materiales y Métodos</b>	<b>39</b>
6.1. Participantes . . . . .	39
6.2. Adquisición de imágenes . . . . .	40
6.3. Estímulos y tarea . . . . .	41
6.4. Métodos de análisis . . . . .	44
6.4.1. Preprocesamiento de imágenes . . . . .	44
<b>7. Resultados</b>	<b>46</b>
7.1. Experimento 1 . . . . .	46

7.2. Experimento 2 . . . . .	49
7.2.1. Resultados conductuales . . . . .	49
7.2.2. Validación del método . . . . .	52
7.2.3. Percepción emocional . . . . .	54
<b>8. Discusión</b>	<b>58</b>
<b>9. Conclusiones</b>	<b>59</b>
<b>Apéndice 1: prueba de la función de verosimilitud de la regresión lineal simple</b>	<b>60</b>
<b>Apéndice 2: prueba de <math>\hat{\theta}</math> por MLE para regresión lineal simple (ecuación normal)</b>	<b>61</b>
<b>10. Apéndice 3: prueba de <math>\hat{\sigma}^2</math> por MLE para regresión lineal simple</b>	<b>62</b>
<b>Apéndice 1: código fuente</b>	<b>63</b>
<b>Referencias</b>	<b>65</b>

# Capítulo 1

## Introducción

### ¿Qué son las emociones? Consideraciones biológicas y psicológicas

Las emociones son reacciones corporales generalizadas y estereotipadas ante estímulos o eventos, ya sean internos o externos, que el organismo evalúa como relevantes en términos de las muchas facetas del éxito evolutivo (Schacter et al. 2011; LeDoux, Damasio 2013). Por ejemplo, el miedo de un gato espeluznado ante un perro — como cualquier instancia de esta emoción — es una respuesta adaptativa<sup>1</sup> forjada en parte en el fenotipo de las especies y en su conducta aprendida para proceder con cautela durante la percepción de peligro. La reacción de una emoción abarca tanto componentes psicológicos como meramente fisiológicos (Schacter et al. 2011; LeDoux, Damasio 2013). En el caso del desafortunado gato se tendrían, por un lado, la relocación súbita de su atención hacia el perro, la sensación misma del miedo como una suerte de compresión de las entrañas de abajo hacia arriba, etc. Por el lado fisiológico podríamos contar la sobreapertura de los párpados, hiperventilación y aceleración del ritmo cardíaco, liberación de hormonas estresantes, piloerección y quizás conductas de congelamiento o huida por parte de la musculatura estriada.

La familia de las emociones es más numerosa que el miedo, si bien las relaciones precisas entre sus miembros son un punto de debate. Por mencionar algunas de las más populares, fenómenos como la alegría, la tristeza, el enojo, el asco y la vergüenza también forman parte de esta familia (Ekman, Davidson 1994). Cada una ameritaría su propia descripción en tanto que fenómeno psicofisiológico, así como su posible explicación funcional (ver “las 4 preguntas de Tinbergen” (Tinbergen 1963) sobre la naturaleza multifacética de las explicaciones en biología).

Entretejidamente a la definición anterior, la palabra “emoción” también se usa para referir específicamente a uno de sus aspectos psicológicos: la sensación característica o experiencia consciente<sup>2</sup> que suele acompañar a los exabruptos fisiológicos anteriormente mencionados (Tsuchiya, Adolphs 2007; Schacter et al. 2011; LeDoux, Damasio 2013). Quienes

<sup>1</sup>Sin ánimos de anatematizar la existencia de vestigios a la par de adaptaciones.

<sup>2</sup>Por “consciencia” me referiré específicamente al significado amplio y rudimentario de conciencia fenoménica: la cualidad de los objetos que sienten, así como el contenido de dichas experiencias subjetivas; en la línea de autores como Christof Koch, David Chalmers, Giulio Tononi y Ned Block.

abogan por una desambiguación proponen hablar de “sentimiento” (feeling) en lugar de “emoción”, y para ellos, “emoción” también rima con “conducta observable” (LeDoux, Damasio 2013). De cualquier manera, desde esta segunda acepción igual de inexorable, dos características emergen como necesarias y suficientes para distinguir la fenomenología de los sentimientos respecto de otros estados mentales (Schacter et al. 2011): la primera es que su intencionalidad está dirigida hacia su respectivo estado somático.<sup>3</sup> La segunda, no sorprendentemente, es que acarrean un valor hedónico; como deja entrever la estrechez entre el sistema de recompensa y el sistema emocional (Berridge, Kringelbach 2008). Es decir, las emociones parecen siempre ser agradables o desagradables, sin lugar a una tercera opción.<sup>4</sup>

Hablar de emociones es hablar de un fenómeno biológico muy básico, bien conservado a través de la filogenia, tanto así que algunos estudiosos sugieren que los rudimentos de las manifestaciones conductuales de las emociones (esencialmente conductas de apertura-acercamiento o rechazo) son trazables a respuestas análogas en todo el árbol de la vida (Damasio, Carvalho 2013); aunque el caso más claro por mucho es el de animales con un sistema nervioso, que es lo que sustenta y coordina sus funciones emocionales. Resulta interesante por lo tanto entender el sustrato de una de las funciones básicas del sistema nervioso; y aunque nadie sano duda poder distinguir la alegría de la tristeza (ya sean las propias o las ajenas), descubrir su naturaleza en actividad nerviosa ha probado ser especialmente dificultoso.

La problemática es patente a partir del abanico de teorías psicológicas que compiten por caracterizarlas en términos conceptuales, materiales y evolutivos. Por un lado se encuentran los modelos que postulan la existencia de emociones “básicas” o atómicas, que pueden variar en intensidad y combinarse para producir emociones más *sui géneris*. Otros modelos las sitúan a todas en un espacio continuo de componentes afectivos más simples (como valencia hedonística y nivel de excitación corporal). En tercer lugar encontramos las teorías de “appraisals”, para las cuales la categorización emocional no se conforma a ningún espacio de dimensión finita, sino que surge de evaluaciones situacionales más finas conforme al contexto (ejemplos: soledad, decepción o culpabilidad en lugar de llana tristeza). Otros más terminarían por argüir que no hay modelos unificadores puesto que se trata de fenómenos corporales muy dispares, y los intentos por agruparlos son poco más que constructos culturales (ver Kragel, LaBar 2016 para una revisión comparativa).

---

<sup>3</sup>A diferencia de, digase, el exquisito olor de las flores; que se corresponde a un objeto externo. Damasio (1996) utiliza el término “somático” en lugar de “corporal” para comunicar la suficiencia de *representaciones* o simulaciones mentales del cuerpo.

<sup>4</sup>A diferencia de, por ejemplo, la experiencia libre de placer o desagrado de un color vívido, o información propioceptiva sobre la posición de las extremidades.

## Emoción y salud

Sin importar cuál sea su organización, sabemos que las predisposiciones en la dinámica afectiva del individuo son una importante marca de la personalidad (Schacter et al. 2011), que como cualquier mecanismo corporal está sujeto a descomposturas. Podría parecer extraño hablar acerca de estados psicofisiológicos de ocurrencia natural y frecuente en relación a la salud; sin embargo su matrimonio no puede ser pasado por alto. En primer lugar, las emociones mismas pueden ser sintomáticas de la salud o la enfermedad (como es el caso del asco), ya que juegan un papel clave en la representación neuronal del estado orgánísmico global (Damasio 1999). Pero más directamente, se reconoce que afecciones a los sistemas de la emoción producen trastornos o psicopatologías (Berrios 1993; Millon, Davis 1996); independientemente de la selección de entidades clínicas y diagnósticas del momento, y del incipiente conocimiento científico de sus etiologías o causas concretas. Se pueden mencionar, a manera de ilustraciones, diversas formas de depresión extendida (trastorno depresivo mayor, distimia), anhedonia, fobias y desórdenes de la ansiedad, así como el trastorno límitrofe o “emocionalmente inestable” de la personalidad.<sup>5</sup>

Aunque de menor prevalencia que las desregulaciones psiquiátricas anteriores, funciones emocionales completas pueden llegar a perderse mediante lesiones, síndromes genéticos, patógenos y otras formas de deterioros neurológicos focalizados a estructuras del sistema emocional. Entre estas estructuras encefálicas dispersas encontramos desde núcleos mesencefálicos (como PGA), diencefálicos (hipotálamo posterior), telencefálicos subcorticales en ganglios basales (NAcc) y en la amígdala,<sup>6</sup> así como partes distantes de la neocorteza de los mamíferos<sup>7</sup> y hasta del cerebelo;<sup>8</sup> resaltando la segregación y complejidad de los sistemas de la emoción. Cabe mencionar que la investigación biomédica en torno a problemas de salud emocional no se limita al *localizacionismo*. Existe un fuerte interés por encontrar explicaciones y biomarcadores clínicos modelando al cerebro como una red compleja de componentes en interacción (Rubinov, Sporns 2010; Chialvo 2010). Por ejemplo, se ha hipotetizado que el síndrome de Capgras — en el que seres queridos son confundidos con impostores — resulta de una pérdida de comunicación entre el sistema de reconocimiento de caras en el lóbulo temporal y el sistema emocional, no obstante que ambos siguen operando bien por separado (Ellis et al. 1997; Ramachandran 1998).

<sup>5</sup>Para listas y descripciones más exhaustivas de uso corriente entre psiquiatras y psicólogos clínicos, se pueden consultar las varias entradas sobre desórdenes depresivos y de ansiedad en el Manual diagnóstico y estadístico de los trastornos mentales (American Psychiatric Association 2013), o las entradas sobre desórdenes afectivos y relacionados con el estrés en el capítulo 10 de la Clasificación internacional de enfermedades (World Health Organization 2004).

<sup>6</sup>Lo que ocasiona el síndrome de Klüver-Bucy (Klüver, Bucy 1937; Marlowe et al. 1975) o la enfermedad de Urbach-Wiethe (Hurleman et al. 2007).

<sup>7</sup>La demencia frontotemporal en su variante conductual es un caso de síndrome neurológico de origen cortical con fuertes repercusiones emocionales (Rohrer et al. 2011).

<sup>8</sup>La faceta afectiva del síndrome cognitivo-afectivo cerebeloso (síndrome de Schmahmann) está ligada a daño perivermiano del lóbulo posterior del cerebelo (Schmahmann, Sherman 1997).

## Emoción y cognición

Lejos de ser un mecanismo aislado — ¡y en contra del sentido común! —, la cognición más superior y racional es y *debe ser* en cierta medida modulada y sazonada por emociones. Veamos un caso de ello: en estado fisiológico la gente usa sus coronadas emocionales como información para nutrir sus creencias sobre la probabilidad de éxito en una actividad (Isen, Patrick 1983). ¿Por qué sería útil y heurístico este sesgo? La respuesta es simple: el mundo causa en gran medida nuestros estados afectivos; por lo tanto los estados afectivos pueden proveer información sucinta — ¡vaya!, de calidad ejecutiva — acerca del mundo (Schwarz, Clore 2003).

También aquí la patología ofrece una oportunidad para poner de manifiesto la relevancia cognitiva de la emocionalidad. La paciente S.M., quien perdió ambas amígdalas debido a la enfermedad de Urbach-Wiethe, mantiene un desempeño ileso en pruebas comunes de conocimientos, reconocimiento, inteligencia fluida y memoria. ¿Qué la lleva entonces a apostar y perder todo su dinero en simulaciones,<sup>9</sup> a mostrar dificultades decidiendo sobre la prioridad de las opciones que se le presentan en su vida cotidiana, y a mostrarse transgresora al punto de la ineptitud interpersonal? El problema no está en su habilidad para razonar, sino en su incapacidad para sentir. La ausencia de instintos como el miedo se traduce en una pésima adopción de estrategias en situaciones emocionalmente cargadas (Damasio 1996; Bechara et al. 2000).

¿Son muestras como éstas las excepciones o la regla? Una vez reconocido que la emoción sobrevive porque otorga ventajas adaptativas, incluidas de índole cognitiva, es natural preguntarse hasta dónde se extienden (o deberían extender) sus dominios. Lo cierto es que un sinfín de conocimientos en aprendizaje por refuerzos (Berridge, Kringelbach 2008), control motor (Kording, Wolpert 2006), inteligencia artificial y robótica (Sutton, Barto 2018), economía (Jevons 1879) y ética<sup>10</sup> (Bentham 1781) solo pueden ser entendidos bajo la luz del valor subjetivo que las emociones esconden; o al menos su operacionalización en forma de utilidad o recompensa. Su relación no es fortuita. Como pieza clave en el rompecabezas de las funciones neurales, la emoción provió los cimientos para algo tan crucial y etéreo como la motivación (Schwarz 1990; Schacter et al. 2011); y por transitividad, sobre éste también descansa la planificación, sin mencionar aspectos de la toma adecuada de decisiones (Schwarz 1990; Bechara et al. 2000), todos ellos característicos de la inteligencia. Parece que el agente racional perfecto cree por evidencia, pero actúa por pasiones; de lo contrario su conducta caería en la quiescencia o en la trivialidad de un arco reflejo.

<sup>9</sup>Como la prueba neuropsicológica del juego de azar de Iowa (Iowa Gambling Task) (Bechara et al. 2001) y la prueba por desliz (Faux Pas) (ver Torralva et al. 2009).

<sup>10</sup>Más allá del utilitarismo de Bentham y Mill, la mayoría de las teorías en ética normativa pertenecientes a la familia consecuencialista están formuladas en términos como “felicidad” y “sufrimiento”.

## Emoción y sociedad

Los fenómenos emocionales no nacen y mueren en el individuo. Existe especial significancia sobre lo que las emociones logran para el bienestar social. Este texto inició por plantear el concepto de emoción como una reacción corporal adaptativa para la supervivencia más inmediata y prosaica, sin embargo esta visión es miope. Es sumamente probable que entre el complejo expresivo y postural de una emoción también encontremos características seleccionadas para la comunicación, y no necesariamente para predisponer al organismo a sortear la situación o aprender de ella (Russell, Dols 1997; Tracy et al. 2015). A continuación se exponen distintos niveles en los cuales la emoción opera como facultad cognitiva social, desde la selección de grupo y sexual hasta las sofisticadas mentalización y empatía.

En etología (el estudio del comportamiento animal) se habla de que una señal del organismo emisor al receptor es efectiva si posee la propiedad de ser ventajosa para ambas partes, en promedio (Donath 2007; Laidre, Johnstone 2013). Por ejemplo, la aparatoso aposematosis de una indefensa rana venenosa, o las costosas estridulaciones a triple *forte* del grillo macho son más que un regalo para disuadir predadores o atraer parejas, respectivamente. Claramente, la rana se beneficia de no ser comida aunque como bocadillo sea letal, así como el beneficio es compartido entre ambos sexos de grillo. Lo mismo estaría pasando con el llanto de un niño, o con la excesiva hilaridad con la que actuaba su madre cuando interese por primera vez en el padre: el llanto también existe para beneficio de la madre, y la jocosidad con la que casi involuntariamente comenzó a seducir al padre —en su momento— fue más que una autoseñalización de estar en un ambiente agradable y confiable.

Más aun, si es difícil fingir o suprimir estas señales (en proporción a la flexibilidad cognitiva de la especie), entonces mayores son los indicios de su efectividad a lo largo de la historia de la evolución. Mentir o prevaricar resta al valor que receptores y emisores honestos obtienen de dichas señales (Donath 2007; Tracy et al. 2015); por lo que resulta sumamente interesante para etólogos y antropólogos notar que, incluso en las especies más inteligentes, la señalización emocional es difícil de fingir (Damasio 1999; Ekman 2003; Porter, Ten Brinke 2008). En efecto, la sociedad humana reconoce a los buenos actores teatrales como ejecutores de proezas dignas de premios.

Es en nada más y nada menos que primates donde la percepción<sup>11</sup> de expresiones emocionales parece haber asido circuitos sumamente especializados, específicamente usando la expresión facial como medio (Tsao et al. 2006). Pocos estímulos son tan informativos como los rostros. Basta una fracción de segundo para que la cara de una persona nos cuente algo

---

<sup>11</sup>Percepción es el conjunto de procesos de interfaz mediante los cuales un sistema cognitivo adquiere información externa y *la interpreta*, volviéndola útil para procesos de orden superior como la categorización y conceptualización, la acción y el control ejecutivo.

acerca de su sexo, edad, etnia; incluso de su foco de atención y por supuesto, su estado anímico (Tsao, Livingstone 2008; Haist, Anzures 2017; Reggev et al. 2020). Quizás ningún otro estímulo induce tantas ilusiones pareidólicas (Liu et al. 2014), y la sofisticación de su reconocimiento — que aparece pocas horas después de nacer y alcanza la cuasi-madurez naturalmente antes del primer año (Haist, Anzures 2017) — solamente es equiparable a habilidades aprendidas arduamente como la lectura.

Se sabe al menos desde tiempos de Hughlings Jackson y Charcot que el cerebro humano puede perder selectivamente la capacidad para identificar rostros. Este desorden neurocognitivo, conocido como ceguera o agnosia de caras (prosopagnosia), depende de regiones corticales especializadas en la vía visual ventral o de reconocimiento de objetos. De particular prominencia es cierta área en los giros fusiformes de la cara ventral de los lóbulos temporales (“Fusiform Face Area”) que responde más fuertemente para caras que para cualquier otro estímulo (Haxby et al. 2000; Said et al. 2010).<sup>12</sup> Mucho antes de la aparición del lenguaje *bona fide* humano, nuestros ancestros tuvieron que leer estas expresiones e interpretarlas como parte de un repertorio de señalización para la vida en pequeñas sociedades de tropa (Fessler, Gervais 2010; Fridlund 2014). Acordemente, las mismas técnicas modernas de neuroimagen muestran evidencia de áreas análogas a FFA en macacos (Tsao et al. 2006), cuyo ancestro en común con humanos data de hace más de 25 millones de años (Lewis 2013). Para rematar, exploraciones más invasivas mediante electrocorticografía en pacientes humanos y arreglos de microelectrodos en el sistema de “parches” del macaco han revelado que las neuronas involucradas responden exclusivamente a rostros (Quiroga et al. 2005; Tsao et al. 2006). Es decir, no solo el agregado de actividad LFP<sup>13</sup> sigue un código para la detección de rostros; el campo receptivo de cada célula individual de alguna manera ya lo está haciendo.<sup>14</sup>

Este sobrevuelo por los extensos territorios de las emociones no podría finalizar sin rozar las cumbres más elevadas del antropocentrismo. Después de la supervivencia, la salud, la cognición y la visita al zoológico... la buena percepción de emociones termina por alimentar los mecanismos de algo tan sofisticado como la teoría de la mente<sup>15</sup> (Mitchell, Phillips 2015) y la empatía (Shamay-Tsoory et al. 2009; Zaki, Ochsner 2012). Siendo las

<sup>12</sup>Bien explican Haxby y coautores (*ibidem*) y Haist, Anzures (2017) que la actividad del surco temporal superior posterior (pSTS) también se ha asociado al procesamiento de los aspectos faciales dinámicos; y tanto FFA como pSTS se alimentan del lóbulo occipital lateral inferior ipsilateral. Por lo tanto, se habla de toda una red de procesamiento de caras.

<sup>13</sup>Local Field Potential: potencial de campo (eléctrico) local. El voltaje del medio extracelular que rodea la membrana de células excitables.

<sup>14</sup>Aunque se han encontrado neuronas individuales cuya actividad correlaciona con la presentación de identidades personales individuales (Quiroga et al. 2005), es más plausible que tanto el reconocimiento de rostros como su individuación en realidad estén siguiendo un código poblacional, como sugiere la evidencia expuesta por Grossman et al. (2019).

<sup>15</sup>Teoría de la mente: atribuir estados mentales a otros. La creencia de que existen otras mentes; en oposición al solipsismo. No debe confundirse con las teorías de lo que la mente es.

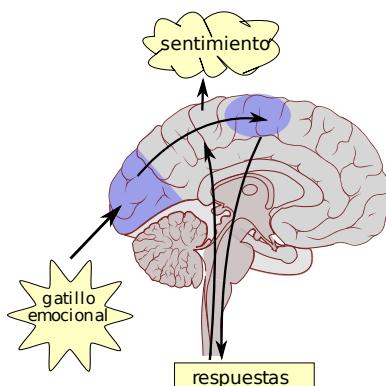
emociones una flagrante muestra de los estados mentales más viscerales, todo el tiempo estamos leyendo las marcas de su expresión en el otro, retroalimentando nuestros propios aparatos cognitivos para poder lidiar con los demás, o bien, empatizar y sentir con ellos. Quizás una explicación de este tipo, un elaborado refriego de reflejos de antaño, subyace controvertidos fenómenos (para las ciencias biológicas) como la adopción o la compasión que otras especies nos merecen (Dawkins 1976; Wilson 2012). Más allá de ser ajenas al humano y lo humano, las emociones son parte y arte de los complejos razonamientos que acarrea la vida en sociedad.

# Capítulo 2

## Antecedentes

### 2.1. Bases neurales de las emociones

El conocimiento de las bases nerviosas de las emociones tuvo inicios tumultuosos y sigue sufriendo cambios y refinamientos. El destacado filósofo y pionero de la psicología, William James, leyó *The Expression of the Emotions in Man and Animals* (Darwin 1872) y encontró inspiración para lanzar abiertamente la pregunta a la comunidad científica: ¿qué es una emoción?, no sólo en términos etológicos y evolutivos como incitó Darwin, sino también neuropsicológicamente hablando (James 1884). La primera propuesta, debida a él y Carl Lange, conforme al conocimiento de la época fue que habría regiones en la corteza para evaluar la información sensorial y otras para controlar efectores de respuesta (James 1884; Cannon 1927). También se decía que las sensaciones o sentimientos resultaban de que el cerebro se percata de los cambios periféricos que él mismo había solicitado. En otras palabras, que no eran consecuencia directa de experimentar los estímulos, sino de experimentar el componente fisiológico de la emoción. Aunque los modelos neurobiológicos han cambiado bastante desde entonces, el ordenamiento de la cadena causal de eventos que llevan a un sentimiento es tema recurrente en las pesadillas de los estudios contemporáneos de la emoción.



Teoría de retroalimentación periférica  
James-Lange, 1884

Figura 2.1: Teoría James-Lange de la emoción. Tanto la evaluación de carga emocional en el estímulo o evento interno, la producción de una respuesta descendente y el sentimiento son la prerrogativa de partes indefinidas de la corteza cerebral. Nótese cómo los sentimientos ocurren tardíamente con la retroalimentación ascendente.

Más tarde, fisiólogos de renombre como Charles Scott Sherrington y Walter B. Cannon mostraron experimentalmente que respuestas emocionales innatas eran posibles en animales completamente decorticados y en otros donde se había desconectado el sistema nervioso autónomico, pero no en ausencia del tálamo e hipotálamo (Sherrington 1900; Cannon 1927). Algo andaba muy mal con la teoría de James-Lange. El nuevo modelo con piedra angular en el diencéfalo, bautizado como de Cannon-Bard, fue complicándose con la adición teórica del giro límbico o cingulado y del hipocampo, según las observaciones de conectividad anatómica de James Papez (Papez 1937); así como el descubrimiento de funciones viscerales (Triarhou 2008) para esta corteza “límbica” o “anillada”, y correlaciones entre la conducta agresiva de la rabia y sus efectos morfométricos en el hipocampo (Eysenck 2004). Paul MacLean extendió el circuito de Papez con más estructuras perihipocampales y de línea media en lo que se conoce como “sistema límbico”<sup>1</sup>, y consolidó la promesa del hipocampo como habitáculo de los sentimientos (MacLean 1949). Esto venía motivado por descubrimientos recientes de que la disección de los lóbulos temporales en monos generaba respuestas emocionales completamente erráticas (Klüver, Bucy 1937).

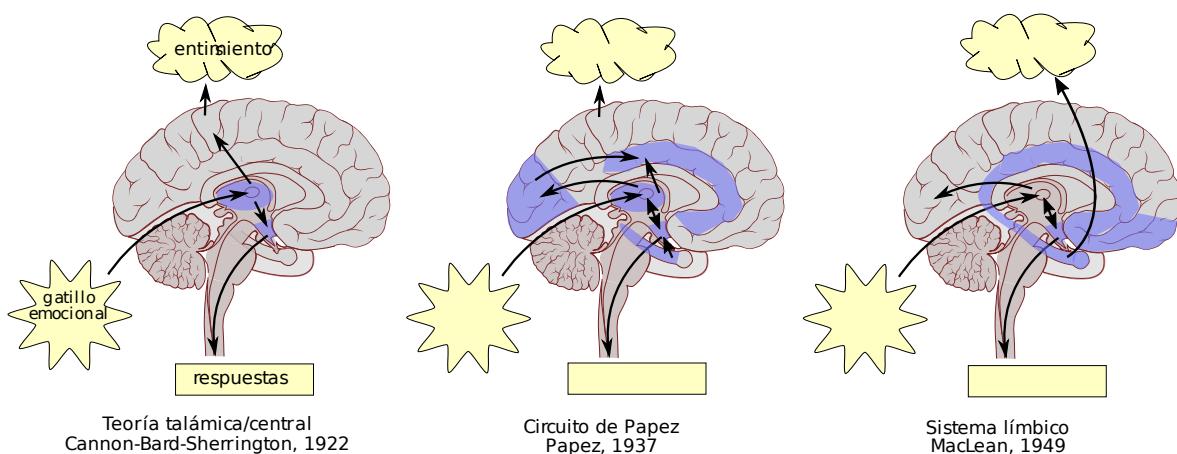


Figura 2.2: Progresión de modelos en el linaje de Cannon-Bard, tomando como punto de partida el descubrimiento de la suficiencia del tálamo y los núcleos del hipotálamo posterior para coordinar enojos innatos. Cannon supo sintetizar los descubrimientos de Sherrington, los de Bard y los propios para proponer una nueva visión *cerebrocentrista* de las emociones: la teoría de los núcleos talámicos o teoría central; que cimbró empíricamente a la de William James. Papez y MacLean harían adiciones parcialmente acertadas.

El sistema límbico sufrió un primer golpe después del famoso paciente H.M. (LeDoux, Damasio 2013), pues se mostró que las funciones del hipocampo, hasta entonces desconocidas, eran completamente distintas. Sin embargo quedaba la incertidumbre de a qué estructura temporal se debían esas anomalías emocionales. Fue el psicólogo Lawrence Weiskrantz, conocido también por el fenómeno de “blindsight”, quien descubrió que se trataba de la amígdala, no del hipocampo (Weiskrantz 1956). Desde entonces la amígdala

<sup>1</sup>Que figuraba prominentemente en su desacreditada teoría de neuroetología evolutiva: el cerebro triúnico (Cesario et al. 2020), de gran aceptación aún entre el público general.

figura como una pieza clave para evaluar la carga emocional de los estímulos y aprender inconscientemente a asociar ciertas situaciones con ciertas respuestas emocionales.

A grandes rasgos, los impulsos tempranos que entran al tálamo son copiados a cortezas sensoriales pero también a la amígdala en cada hemisferio, especialmente sus núcleos laterales (LeDoux, Damasio 2013). Éstos últimos se especializan en detectar eventos emocionalmente relevantes. Una larga historia de investigación en condicionamiento iniciada por Weiskrantz ha mostrado que la amígdala debe estar forzosamente presente para aprender a asociar implícitamente conductas y estímulos condicionados con el miedo, aunque por sí sola no es suficiente (LeDoux, Damasio 2013). Por ejemplo: cuando un niño aprende a no introducir sus dedos en una toma de corriente, sus amígdalas hacen posible que la respuesta de miedo se “copie” de la experiencia intrínsecamente aversiva de una descarga eléctrica a la mera observación de la toma de corriente (siguiendo una lógica de condicionamiento clásico). Luego el núcleo central de cada amígdala proyecta a lugares como el hipotálamo y la sustancia gris periacueductal (PGA) del mesencéfalo, que a su vez activan, respectivamente, actividad humorral estresante en el eje hipofisiario-adrenal y conductas motoras automáticas (LeDoux, Damasio 2013).

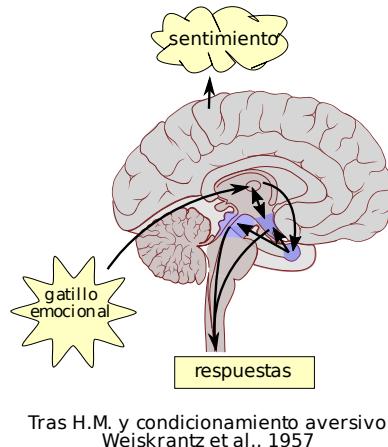


Figura 2.3: La amígdala se incorpora como pieza irrenunciable del sistema de las emociones.

Hoy en día se reconoce ampliamente que algunas zonas de la corteza también juegan un papel en el sistema emocional. De alguna manera siempre fue sospechado, gracias a casos como el de Phineas Gage y demás pacientes con lesiones que llevaron a conductas sociopáticas (LeDoux, Damasio 2013). Las regiones en cuestión son la corteza prefrontal ventral y medial alrededor de los giros rectos<sup>2</sup> (Bechara et al. 2000), la corteza del cíngulo anterior más ventral (vACC) (Bush et al. 2000; Allman et al. 2001), contigua a los giros rectos y ya mencionada como parte del sistema límbico; además de los lóbulos de las

<sup>2</sup>Corteza prefrontal ventromedial (vmPFC), aunque a veces el énfasis se hace más lateral, hacia la parte orbitofrontal (OFC).

ínsulas en sus porciones más anteriores y ventrales<sup>3</sup> (vAIC) (Gu et al. 2013).

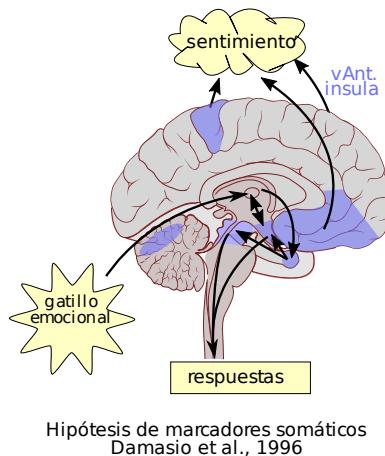


Figura 2.4: El entendimiento contemporáneo del sistema emocional abarca varios componentes corticales e incluso cerebelosos. En imagen se muestran áreas sugeridas por Damasio, incluyendo la corteza somatosensorial primaria (LeDoux, Damasio 2013).

En este punto se han acumulado cientos de estudios sobre los correlatos neurales de las emociones, sin embargo no han podido responderse preguntas un poco más refinadas, como cuáles son las marcas ya no de las emociones como un todo, sino de cada emoción en particular. La localización y codificación de emociones tanto en el sistema nervioso central como en el periférico permanece un problema abierto (Kragel, LaBar 2016; Celeghin et al. 2017). Una pregunta fundamental es si existen patrones en actividad neuronal suficientemente específicos que correspondan con concepciones vernáculas y académicas de las emociones. Además, ¿qué tan generalizables son esos correlatos entre individuos y especies? Esta última cuestión tiene acaloradas implicaciones para la comprensión de la aparición de estados afectivos, ya sea en términos de evolución genética o cultural.

Una técnica o familia de técnicas que resulta especialmente útil para mapear funciones cerebrales tan distribuidas (como las emociones) es la neuroimagen funcional.<sup>4</sup> Desafortunadamente, los metanálisis y revisiones de la literatura no muestran una tendencia en la respuesta a las preguntas anteriores. Phan et al. (2002) analizaron los mapas de activación de 55 estudios de PET<sup>5</sup> y fMRI<sup>6</sup>, y encontraron evidencia parcial para la existencia de correlatos neuroanatómicos consistentes a la ocurrencia de distintos tipos de emocio-

<sup>3</sup>Curiosamente, tanto la ínsula anterior como el cíngulo anterior mantienen la misma subdivisión durante su fuerte interacción en funciones emocionales y de monitoreo/autorregulación. Mientras que el cíngulo anterior dorsal y la ínsula anterior dorsal se han asociado a la red de saliencia en estudios en estado de reposo, sus contrapartes ventrales están estrechamente ligadas a la emoción.

<sup>4</sup>El uso de imágenes cerebrales para medir algún aspecto de la actividad cerebral, y que abarca técnicas tan diversas como el EEG (electroencefalograma, al mismo tiempo una técnica electrofisiológica), MEG (magnetoencefalograma, estrechamente relacionada al EEG), fNIRS (spectroscopía cercana al infrarrojo), SPECT (tomografía computarizada por emisión de fotón único), ultrasonido funcional, PET y fMRI.

<sup>5</sup>Positron Emission Tomography: tomografía por emisión de positrones

<sup>6</sup>Functional Magnetic Resonance Imaging: imagenología de resonancia magnética funcional.

nes básicas. Murphy et al. (2003) incrementaron el tamaño de muestra a 106 estudios y mantuvieron la conclusión general respecto del metanálisis anterior, aunque los correlatos no fueron suficientemente similares. La relativa debilidad de los resultados ha sido usada como evidencia de hipótesis alternativas; a saber, teorías dimensionales y de actos conceptuales (emociones como construcciones de afectos, Barrett 2006). Barrett, Wager (2006) examinaron 161 estudios y encontraron correlatos exclusivos (mas no específicos) para miedo, tristeza y asco. La felicidad y el enojo no arrojaron correlatos ni consistentes ni específicos. Vytal, Hamann (2010) utilizaron un análisis con mayor sensibilidad espacial para contrastar cinco emociones básicas provenientes de 83 estudios. Concluyeron que aunque existen *clusters* característicos de activación para cada una de ellas, no es evidencia suficiente para descartar otros modelos de representación emocional en el cerebro. Lindquist et al. (2012) analizaron 91 estudios tanto de experiencia de emociones como de percepción emocional. Definieron evidencia a favor de emociones básicas como actividad consistente en áreas específicas selectivas a una única categoría. Se encontró consistencia y selectividad para las cinco categorías bajo consideración, mas no especificidad funcional. Es decir, las escurridizas emociones básicas tienen correlatos distinguibles entre sí, aunque las mismas regiones cerebrales están vinculadas a otras funciones psicológicas extraemocionales.

Quizás la manera más efectiva de comunicar semejante discurrencia sea con el siguiente contraste entre dos citas a los resúmenes de los dos metanálisis más recientes:

“Each of the emotions examined [...] was characterized by consistent neural correlates across studies, as defined by reliable correlations with regional brain activations. In addition, the activation patterns associated with each emotion were discrete (discriminable from the other emotions [...])”. Vytal, Hamann (2010).

[W]e found little evidence that discrete emotion categories can be consistently and specifically localized to distinct brain regions. Instead, we found evidence that is consistent with a psychological constructionist approach”. Lindquist et al. (2012).

No pueden ambos estar en lo correcto.

El consenso disponible (y casi unánime) es que la búsqueda de estructuras específicas a un tipo de emoción está superada (Murphy et al. 2003; Kober et al. 2008; Lindquist et al. 2012). Estructuras cerebrales que parecían tener una correspondencia uno-a-uno con alguna emoción básica, como la amígdala (miedo) y la ínsula (asco) (Calder et al. 2001), ahora son acreedoras a funciones más elusivas o generales (Sander et al. 2003). Además,

si la función emocional tiene biomarcadores reproducibles, su decodificación<sup>7</sup> dependerá de la actividad conjunta de redes funcionales distribuidas y traslapadas (Hamann 2012; Lindquist, Barrett 2012; Kragel, LaBar 2014; Celeghin et al. 2017; aunque comparativamente no se han realizado tantos estudios usando técnicas de electrofisiología, la revisión realizada por Guillory, Bujarski 2014 concuerda con esta conclusión).

En este sentido, el campo de neurociencia afectiva aunado a técnicas como fMRI ha promovido ampliamente la incorporación de nuevos métodos de análisis, especialmente métodos multivariados,<sup>8</sup> puesto que podrían ofrecer mayor sensibilidad para responder la pregunta de los correlatos de las escurridizas emociones. Antes de poder contrastar las bondades y operaciones de éstos contra el análisis clásico se precisará recordar brevemente el origen de las señales más usualmente buscadas en fMRI: las señales BOLD o “dependientes del nivel de oxígeno en la sangre”.

## 2.2. Resonancia magnética y el efecto BOLD

Cuando de registrar actividad cerebral con la mayor envergadura y resolución espacial se trata, la técnica más conspicua actualmente es la resonancia magnética nuclear funcional. Escáneres modernos de este tipo son capaces de registrar casi simultáneamente una forma de metabolismo aeróbico en todas las partes de una malla virtual, que bien podría cubrir la totalidad del encéfalo. La resolución temporal de estas señales, conocidas como “BOLD”, difícilmente puede ir por debajo de 1 ó 2 segundos. Sin embargo, está correlacionada con la actividad de potencial de campo local de las neuronas (Logothetis et al. 2001). Es una técnica idónea para estudios en los que redes funcionales asociadas a tareas primero deben ser identificadas o refinadas — *in vivo* y sin invasividad — allanando el terreno para técnicas de medición más finas que intenten establecer la circuitería y química sináptica (Logothetis 2008). Estos párrafos pretenden esbozar los fundamentos físicos de la resonancia magnética nuclear en general, y su aplicación al estudio de la función cerebral como representante tácito de muchas otras aplicaciones que se le han encontrado tan solo en el campo de las imágenes médicas.

El efecto BOLD se ha convertido en la forma más común de hacer fMRI. Éste fue descubierto (o quizás redescubierto) por Seiji Ogawa y colaboradores en una serie de estudios al inicio de la década de 1990 (Ogawa, T.-M. Lee, Nayak, et al. 1990; Ogawa, Lee 1990;

---

<sup>7</sup>(De)codificación es un término importado de la teoría de la información para referirse a la transformación de la información de una representación a otra. En neurociencia computacional, codificación refiere específicamente al conocimiento y aplicación de las funciones matemáticas que convierten estímulos o entradas sinápticas a su respectiva actividad de salida (ej. potenciales de acción). Decodificación es el proceso inverso: inferir entradas a partir de salidas.

<sup>8</sup>Se usa el calificativo “multivariado” en el sentido estadístico y matemático: que trabaja con funciones de múltiples variables.

Ogawa, T.-M. Lee, Kay, et al. 1990; Ogawa et al. 1992). La señal BOLD es una medida indirecta de los incrementos y disminuciones de actividad neuronal. Ante una fuerte demanda metabólica de las neuronas, mecanismos diversos de acoplamiento neurovascular responden provocando vasodilatación para compensar la extracción del oxígeno en los eritrocitos que van pasando por los capilares aledaños (ver figura 2.5). Más incluso, la compensación es una *sobrecompensación* (Uludag et al. 2005); por lo que el balance neto conforme aumenta la tasa de disparo de potenciales de acción (junto con otros fenómenos celulares) es hacia la hiperemia e hiperoxia. Décadas atrás, el famoso químico Linus Pauling había descubierto que la molécula de la hemoglobina presenta una susceptibilidad magnética distinta dependiendo de si los dominios con hierro portan o no oxígeno (Pauling, Coryell 1936). Se puede imaginar que mientras que la versión oxigenada de la hemoglobina preserva el flujo del campo magnético en el que está inmerso el tejido, su contraparte desoxigenada lo deforma. Es entonces cuando la prestación del fenómeno a ser cuantificado mediante resonancia magnética se vuelve natural y patente: la señal BOLD como fluctuación electromagnética es una función de la tasa metabólica de oxígeno en la sangre local y, sobretodo, del flujo sanguíneo (Uludag et al. 2005).

Todas las aplicaciones usuales de resonancia magnética nuclear buscan excitar el *espín* de algún isótopo particular, usualmente hidrógeno protio ( $^1H$ ) en el ambiente imagenológico, de gran abundancia en las moléculas de agua y biomoléculas. Los espines yacen en un estado basal de energía, alineados preponderantemente como brújulas en dirección del fuerte campo del magneto externo, hasta que un pulso electromagnético sintonizado a su frecuencia natural de resonancia los hace pasar a otro estado excitado en dirección opuesta. Según los protones van relajándose al estado inicial, la energía electromagnética es disipada de vuelta; no obstante, aquellos que se encuentran en un campo magnético diezmado se desfasan más rápidamente en su giro en precesión; por lo que la señal transversal al campo magnético principal, tal y como es obtenida en una región del tejido, es menor cuando la actividad neuronal y el flujo de sangre oxigenada es menor. Ésta es la resonancia magnética nuclear que subyace aplicaciones como el efecto BOLD, condensada a una cáscara de nuez. ¿Pero qué significaría todo eso con algo más de detalle?

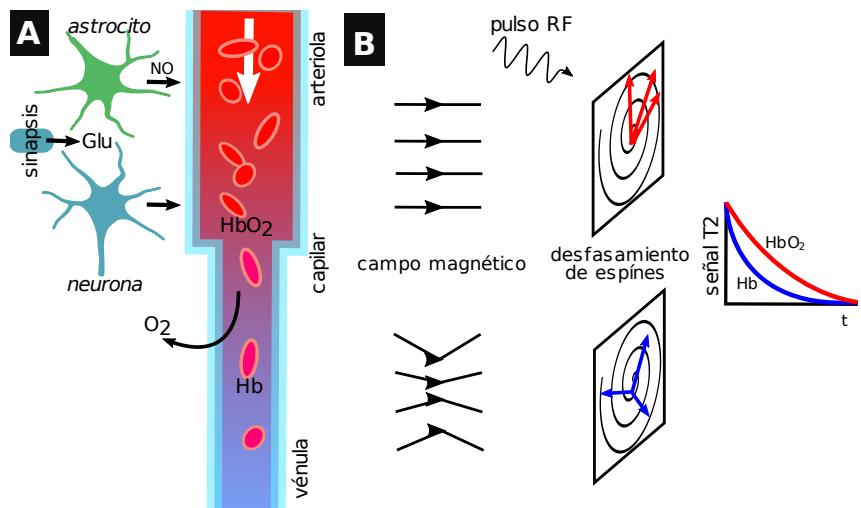


Figura 2.5: Fisiología y física básicas de la señal BOLD. **A:** unión neuro-astrocítica-vascular provoca vasodilatación además de consumir una fracción del oxígeno hematogénico del capilar. **B:** efecto del fenómeno fisiológico sobre el campo magnético, y cómo es aprovechado para medir una señal dependiente de la actividad sanguínea. Créditos.<sup>9</sup>

El espín ( $S$ ) es una de las propiedades fundamentales de la materia según la física contemporánea, al igual que la carga electromagnética y la masa. El espín es una forma de ímpetu angular o cantidad de movimiento angular (también llamado “momento angular”), como el giro de una rueda o la traslación de los planetas a través de su órbita. A diferencia de esos momentos angulares, el espín se distingue por ser un movimiento en ausencia de cambio de posición; es decir, un giro “sobre su propio eje”, como la rotación de la Tierra. Para sorpresa de los físicos de la década de 1920, el espín resultó ser más que una descripción sumaria de un desplazamiento cíclico de varias partículas (como las que conforman la Tierra en rotación), sino que se encuentra de forma muy sofisticada en las escalas más microscópicas e indivisibles del Universo. Este espín fundamental fue propuesto en primer lugar como aditamento a la mecánica cuántica básica del electrón (Pauli 1924; Uhlenbeck, Goudsmit 1925; Pauli 1927), poniendo fin a los misteriosos resultados de experimentos como el de Stern-Gerlach (Gerlach, Stern 1922b; Gerlach, Stern 1922a) y el efecto Zeeman (Preston 1898).

Como propiedad intrínseca a cada partícula fundamental o campo cuántico, el espín expresa la magnitud de algo abstracto, aunque concebible en un primer abordaje como la rotación que sufre un punto en el espacio sobre sí mismo antes de regresar a la configuración de inicio.<sup>10</sup> Más propiamente, se trata de una transformación en un espacio puramente matemático, interno a la partícula o asociado a ésta, sin tratarse de una rotación de la partícula misma (Pauli 1927).

<sup>9</sup>Neuron Silhouette (Scidraw 2020).

<sup>10</sup>Ralph Kroning ofreció en 1925 dicha interpretación física para la primera formulación de Pauli (1924); pero éste último la rechazó porque la rotación sería más rápida que  $c$ .

Por lo tanto el espín se concibe como cantidad vectorial: además de magnitud, posee dirección o signo que está sujeto a cambios, si bien son cambios discretos. Sistemas de partículas como bariones y núcleos atómicos enteros también pueden acumular un espín distinto de 0, siempre que el número de neutrones o protones sea impar. El espín como momento angular se calcula a partir del número cuántico de espín ( $s$ ), magnitud adimensional que solo toma valores múltiplos de  $\frac{1}{2}$ , según las indagaciones experimentales que se han hecho desde Stern y Gerlach. Este hecho empírico puede ser difícil de asimilar, sin embargo dio pauta a tan peculiar desarrollo matemático. Por ejemplo, los únicos valores posibles para el  $s$  del  $^1H$  congruentes con sus efectos medibles como momento angular y momento magnético son  $+\frac{1}{2}$  y  $-\frac{1}{2}$ . Obtener  $S$  a partir de  $s$  simplemente conlleva multiplicar por  $\hbar$ : la constante reducida de Planck.

Como en cualquier esquema de mecánica cuántica, el estado del sistema existe en un espacio de Hilbert con tantas dimensiones como estados, de manera que el  $S$  del  $^1H$  —antes de ser medido— puede estar en una superposición de ambas posibilidades. Esto se modela como el vector unitario  $\mathbf{s}$  (o el *ket*  $|S\rangle$  siguiendo notación de Dirac). Para el caso no relativista de partículas con espín  $\frac{1}{2}$  se tiene:

$$|S\rangle = \mathbf{s} = \begin{pmatrix} a \\ b \end{pmatrix} \quad a, b \in \mathbb{C}; \quad |a|^2 + |b|^2 = 1 \quad (2.1)$$

El vector complejo  $\mathbf{s}$  también es conocido como un *espinor*.<sup>11</sup> Siguiendo la regla de Born,  $|a|^2$  y  $|b|^2$  son las probabilidades de observar  $S$  como  $+\hbar/2$  ó  $-\hbar/2$ .

Es claro que elegir entre  $a$  o  $b$  para representar el signo es meramente arbitrario, por lo que se necesita un método de cambio de coordenadas para que distintos experimentadores traduzcan sus resultados, y para seguir la pista dinámica del sistema conforme evolucione según la ecuación de Schrödinger. Sin embargo, al tratarse de vectores complejos, no es posible traducir el estado cuántico  $\mathbf{s}$  entre rotaciones de los ejes usando el grupo de matrices  $SO(3)$  para espacio euclíadiano 3D. Ejemplo insólito de ello es que si uno midiera el espín de un átomo de hidrógeno estático, llamaría a eso  $|\uparrow\rangle$ , lo rotaría  $360^\circ$  y luego volviera a medir; ¡el espín observado de hecho sería el opuesto a  $|\uparrow\rangle$ !

Mientras tanto, en espacio 3D la medición de  $S$  requiere de una base con tres observables; en coordenadas cartesianas:  $S_x$ ,  $S_y$  y  $S_z$ . Por lo tanto, para rotar un espín  $\frac{1}{2}$  se precisa una base de operadores lineales de  $2 \times 2$ , que llamaremos  $\Sigma_x$ ,  $\Sigma_y$  y  $\Sigma_z$  (las matrices de Pauli). Como todo operador en mecánica cuántica, además deben ser autoadjuntos (matrices Hermitias) y  $\det(\Sigma) = 1$  para preservar la probabilidad o norma de  $\mathbf{s}$ , y sus

---

<sup>11</sup>Otra representación popular equivalente para sistemas de dos niveles (qubits) es la superficie de la esfera de Bloch.

*eigenvectores* (también llamados autovectores o vectores propios) deberían ser los únicos estados posibles (en este caso:  $(+\hbar/2, 0)^T$  y  $(0, -\hbar/2)^T$ ); de modo que existe la certeza de que el operador jamás convertirá estados posibles en estados imposibles.

Aquí es donde el álgebra abstracta y la teoría de grupos no abelianos salvan el día, encontrando una simetría homomórfica de tipo sobreyectivo entre el grupo de Lie  $SO(3)$  (“special orthogonal”) para componer rotaciones continuas en espacio físico, con base ortonormal canónica:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

y sus correspondientes matrices en el grupo  $SU(2)$  (“special unitary”):

$$\boldsymbol{\Sigma}_x = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}, \boldsymbol{\Sigma}_y = \begin{pmatrix} 0 & -i/2 \\ i/2 & 0 \end{pmatrix}, \boldsymbol{\Sigma}_z = \begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix}. \quad (2.2)$$

Aunque el principio de incertidumbre impide asignar valores definidos y simultáneos a cada componente, representación vectorial sigue siendo de gran utilidad cuando se toma en cuenta que el espín de la partícula en mancuerna con su carga producen un *momento de dipolo magnético de espín*,  $\boldsymbol{\mu}$ , como si se tratase de una brújula que experimentará realineación o torque conforme al campo magnético:

$$\boldsymbol{\mu}_s = \gamma \mathbf{s} \quad (2.3)$$

Donde  $\gamma$ , llamada “razón giromagnética”, es una constante particular a cada partícula o sistema ya que está dada por su masa, su carga y su factor-g, según  $\gamma = gq/2m$ .

Cuando el campo magnético es nulo, ambos estados de  $s$  ( $+\frac{1}{2}$  y  $-\frac{1}{2}$ ) son equiprobables o de misma energía, pero en presencia de un campo magnético  $\mathbf{b}_0$ , la energía de cada estado será el producto interior:

$$E = -\mathbf{b}_0 \cdot \boldsymbol{\mu}_s \quad (2.4)$$

Si el campo es despreciable excepto en una dirección (llámese  $z$ ), como en el caso de un solenoide, la expresión (2.4) se reduce a  $E = -b_{0z}\gamma\hbar s_z$ . Es decir: que la energía para cada estado es una función lineal de la magnitud del campo, cambiando la pendiente según el signo de  $s_z$ . Lo que esto significa para el tejido o muestra rodeado por un electroimán es

que el espín de todas sus partículas con un momento magnético se “realinea” en dirección de  $\mathbf{b}_0$ , o la exactamente opuesta, pero no en la misma proporción.<sup>12</sup> La ligera preferencia por ocupar estados de menor energía se traduce en un *vector neto de magnetización de espín* distinto de cero, llamado  $\mathbf{m}$ , para el cual el análisis puede proseguir sin hacer alusión a la mecánica cuántica subyacente.

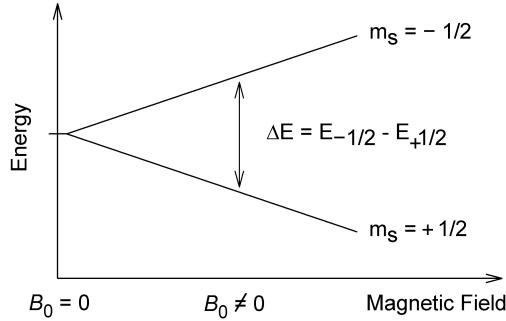


Figura 2.6: Ruptura de los niveles de energía del momento magnético de espín de un protón en presencia de un campo magnético externo. Imagen donada al dominio público por J. Bancroft Brown.

Si ese fuera el fin de la historia no habría resonancia ni aplicaciones interesantes. Sin embargo resulta que, narrado de manera un tanto informal,  $\mathbf{m}$  no sólo gira incesantemente sobre su propio eje debido al espín; y no sólo rota momentáneamente un poco más con tal de volverse paralelo a  $\mathbf{b}_0$ .  $\mathbf{m}$  además sufre precesión: una reorientación cíclica alrededor del eje de  $\mathbf{b}_0$ , similar a los tambaleos de un trompo desalineado en su giro. El movimiento de  $\mathbf{m}$  llevará por nombre y apellido “*precesión de Larmor*.<sup>13</sup> La frecuencia de la precesión de Larmor sólo depende de la razón giromagnética y la magnitud de  $\mathbf{b}_0$ :

$$\nu = \frac{\omega}{2\pi} = \frac{\Delta E}{h} = \frac{-\gamma|\mathbf{b}_0|}{2\pi} \quad (2.5)$$

Mera sustitución de valores permite conocer cuál sería el límite de intensidad del campo magnético antes de tener que exponer la muestra a ondas ionizantes para lograr la precesión.

La resonancia propiamente es una característica matemática de toda ecuación diferencial con soluciones ondulatorias a la cual se ha añadido un término inhomogéneo (es decir, que no depende de la incógnita). Lo que sea que esté representando el término inhomogéneo sirve como suministro de energía al sistema; y cuando esto ocurre en sincronía a la

<sup>12</sup>La distribución de Boltzmann nos dice que la probabilidad de encontrar  $\mu_S$  en uno u otro estado energético es (a temperatura  $T$  y usando la constante homónima  $k$ ):

$$P(E) = \frac{1}{\sum e^{-E_i/kT}} e^{-E/kT}$$

<sup>13</sup>En una verdadera explicación cuántica se hablaría de oscilaciones de Rabi.

ondulación natural, la energía es aprovechada para excitarla.

Algo sobresaliente es que la precesión de Larmor existe por la mera presencia del campo magnético estático, pero su amplitud es contingente a que un campo  $\mathbf{b}_1(t)$ , oscilando a la frecuencia adecuada en el plano  $xy$ , se monte en superposición con  $\mathbf{b}_0$ . Más aún, el pulso resonante sincroniza la precesión. Entre más duradera la pulsación de  $\mathbf{b}_1(t)$ , mayor la apertura del cono que  $\mathbf{m}$  forma mientras precesa. Inclusive podría pasar de precesar en un cono alrededor de la dirección  $z$  a girar como manecilla sobre el plano ortogonal  $xy$ ; con el origen del vector siempre pinchado al origen de este marco de referencia imaginario. Un pulso con el doble de duración ( $t_p$ ) alimentaría suficiente energía para invertir la precesión en la dirección  $z$  opuesta. Así, los pulsos de radiofrecuencia pueden ser identificados según su “flip angle”: de  $90^\circ$ ,  $180^\circ$ , etc. Si la amplitud ( $A$ ) del oscilador externo  $b_1(t) = A\sin(\omega t + \varphi)$  es constante (pulso “cuadrado”), entonces el ángulo obtenido se aproxima a:

$$\angle(\mathbf{b}_{0z}, \mathbf{m}) \approx \gamma At_p \quad (2.6)$$

Una vez que dicho pulso resonante haya cesado, el conjunto de espines irá perdiendo la energía absorbida debido a la sarta de interacciones que siguen ocurriendo entre ellos mismos y con el medio. El vector  $\mathbf{m}$  precesa en espiral de regreso al estado base. Imagine que el sistema acaba de ser excitado con un pulso de resonancia de  $90^\circ$ . Mientras que  $\mathbf{m}$  va perdiendo su componente en el plano transversal “ $xy$ ” a su regreso (relajación transversal), el eje  $z$  lo ve recuperarse (relajación longitudinal). Dichos movimientos de relajación son descritos clásicamente por el sistema de ecuaciones diferenciales de Bloch, con parámetros de tiempo  $T1$  (longitudinal) y  $T2$  (transversal), los cuales son característicos de cada compuesto o tejido. De aquí que la señal electromagnética devuelta registrada por una bobina-antena pueda ser más intensa en una zona del tejido que en otra, o en dos momentos distintos para la misma zona de tejido como sucede con el efecto BOLD.

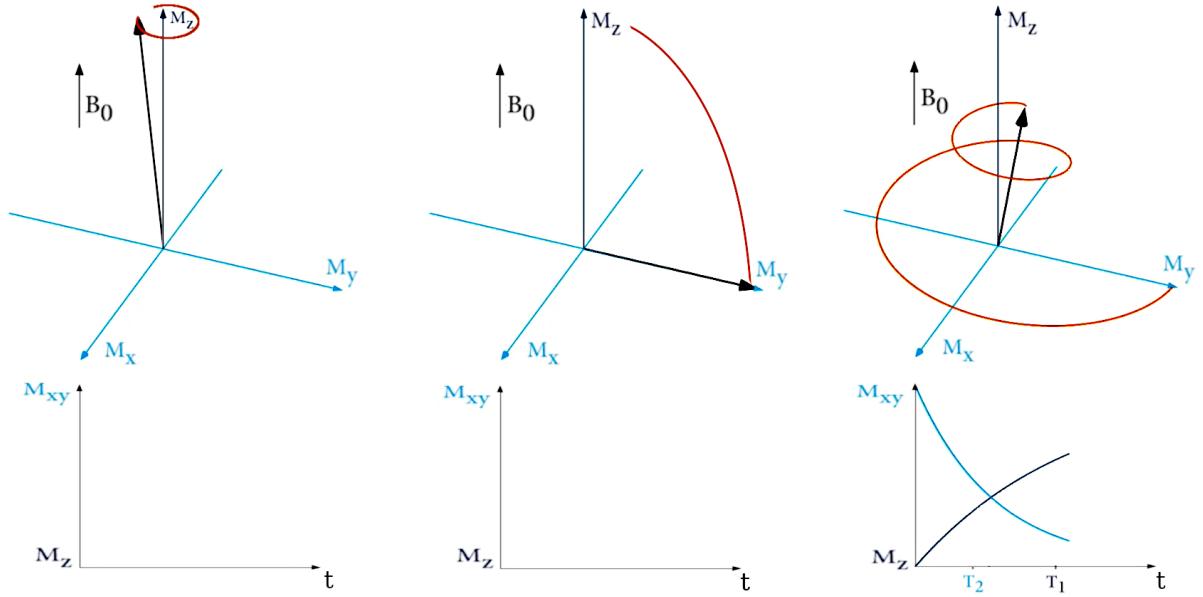


Figura 2.7: Columna izquierda:  $\mathbf{m}$  en precesión. Centro:  $\mathbf{m}$  tras aplicar un pulso resonante de  $90^\circ$  en  $b_1$ . Derecha: relajación de vuelta al estado de inicio. Recuérdese que  $\mathbf{m}$  es la suma de varios momentos magnéticos de espines individuales. La precesión inicial no tiene por qué ser coherente como en el diagrama, pero una vez “tumbados” al plano  $xy$ , todos los espines entran en fase. La velocidad de desfasamiento es crucial para el efecto BOLD.

Se espera que el lector ahora tenga elementos suficientes para lograr una lectura auto-contenida del trabajo, sobre todo durante la descripción del protocolo de imágenes en la sección de *Métodos*. Sin embargo, cabe señalar que no se ha dicho nada al respecto del diseño de secuencias de pulsos para obtener uno u otro tipo de imagen (pesada a  $T_1$ , pesada a  $T_2$ , funcional, pesada a difusión, etc.), ni de cómo se excita selectivamente una rebanada de la muestra, y cómo se reconstruye la imagen a partir del simple muestreo de señales unidimensionales que llegan a la antena y a partir de la transformada inversa de Fourier.

## 2.3. Análisis estadístico de patrones de actividad cerebral

### 2.3.1. Análisis clásico

Es necesario visualizar en conjunto todas las señales de actividad cerebral para entender en qué difiere el análisis clásico respecto del análisis de patrones multivariados. El argumento es una abstracción matemática, y por lo tanto no se limita a considerar exclusivamente los voxels en una secuencia de fMRI y su señal BOLD. Los mismos métodos de análisis son aplicables a registros con electrodos, imágenes de microscopía de doble fotón de calcio

o cualquier otra forma de dinámica cerebral, fisiológica o de cualquier otro campo de estudio.

Friston et al. (1994) introdujeron al mundo de la neuroimagen funcional la forma de análisis que terminaría convirtiéndose en estándar, en un exitoso intento por unificar las distintas herramientas estadísticas lineales que iban aplicándose al incipiente tipo de datos (por ejemplo, Friston et al. 1991; Worsley et al. 1992), y que pretendían otorgar rigor y madurez a las (a veces) abyertas interpretaciones y conclusiones de los estudios.

El análisis clásico busca ajustar un modelo de regresión lineal entre las muestras de datos provenientes de dos o más series de tiempo: la actividad fisiológica y cómo se presentaron los estímulos o cómo aconteció la tarea. El afán es encontrar voxels o pequeñas regiones cuyas señales fisiológicas por sí solas se correlacionan altamente con las variables manipuladas experimentalmente. Aunque el modelo de regresión suele incorporar varias covariables predictoras<sup>14</sup> con información del movimiento del sujeto y otras formas de ruido (para reducir la inflación del verdadero efecto experimental), no se acostumbra incluir múltiples señales de actividad cerebral. De hecho, no se incluye *ninguna*. En lugar de fungir como predictor del experimento, la serie de tiempo de una región cerebral es la variable a ser explicada mediante una combinación lineal o suma ponderada de la variable experimental y covariables. Luego el algoritmo se repite por separado para modelar la señal cerebral en otra zona. Por esta razón, y un tanto erradamente, el análisis de regresión clásico también se denomina “univariado masivo”.

La forma más general de análisis por regresión lineal es el modelo lineal generalizado de efectos mixtos (GLMM), que en notación matricial<sup>15</sup> reza:

$$\mathbf{Y} = g(\mathbf{X}\boldsymbol{\Theta} + \mathbf{Z}\mathbf{U}) + \boldsymbol{\varepsilon} \quad (2.7)$$

Donde  $\mathbf{Y}$  es la matriz de vectores-columna de series de tiempo a explicar (las observaciones de variables dependientes),  $\mathbf{X}$  y  $\mathbf{Z}$  son matrices de diseño con vectores-columna de series de tiempo de regresores (variables independientes); y  $\boldsymbol{\Theta}$  y  $\mathbf{U}$  son las matrices de parámetros a estimar (efectos fijos y efectos aleatorios, respectivamente).  $g$  es una función de activación arbitraria, no necesariamente lineal, pero que se aplica elemento a elemento para preservar las dimensiones de su matriz argumento.  $\boldsymbol{\varepsilon}$  es el error o ruido, modelado como una variable aleatoria que puede pertenecer a cualquier distribución de la familia exponencial.

Si se elige la no-linealidad  $g$  como la simple función identidad y se asume ruido gaussiano,

---

<sup>14</sup>También llamados “regresores” o simplemente “variables independientes”.

<sup>15</sup>O más generalmente, para operadores lineales de espacios vectoriales arbitrarios, discretos como matrices o continuos.

el modelo se reduce al modelo lineal de efectos mixtos (LME,  $\hat{\mathbf{Y}} = \mathbf{X}\Theta + \mathbf{ZU}$ ;  $\mathcal{E} \sim \mathcal{N}(0, \Sigma)$ ). Si se omiten los efectos aleatorios, se obtiene un modelo lineal generalizado (GLM,<sup>16</sup>  $\hat{\mathbf{Y}} = g(\mathbf{X}\Theta)$ ). En la intersección entre ambas simplificaciones se encuentra el modelo lineal general (GLM también,<sup>17</sup>  $\hat{\mathbf{Y}} = \mathbf{X}\Theta$ ;  $\mathcal{E} \sim \mathcal{N}(0, \Sigma)$ ).

El marco teórico del modelo lineal generalizado ya es suficientemente maleable como para capturar diversos tipos de variable de respuesta (booleana, categórica, entera, real, etc.) (Nelder, Wedderburn 1972). Por ejemplo, los GLMs son modelos descriptivos populares de la tasa de disparo de una neurona. Dadas las series de tiempo de potenciales de neuronas presinápticas y las observaciones de los potenciales de acción de la neurona de interés, GLM aprende los pesos que debería poseer cada sinapsis con el afán de que su combinación produzca la respuesta observada (integración espacial de la información). Con algunos trucos también podría aprender el filtro de convolución que determina la contribución de cada potencial de entrada en el pasado (des)polarizando la neurona en el presente (integración temporal).

De cualquier manera, la probabilidad empírica de observar cierto número de espigas en la variable respuesta por unidad de tiempo no es simétrica: no tiene sentido hablar de tasas de disparo negativas, y hay un sesgo a disparar apenas terminado el periodo refractario, por lo que el GLM se adapta con una distribución de probabilidad de Poisson o una exponencial para modelar este tipo de aleatoriedad de la respuesta. También se introduce una función  $g$  que acomode este cambio, respecto de cuando se usa ruido gaussiano para la regresión lineal múltiple.

Cuadro 2.1: Algunas variantes comunes de modelos lineales generalizados (GLMs) y sus peculiaridades matemáticas.

Tipo de GLM	$y$	$\mathcal{E}$	$g$
regresión lineal	$\mathbb{R}$	$\mathcal{N}(\mu, \sigma^2)$	$\mathbf{X}\Theta$ (identidad)
regresión Poisson	$\mathbb{N}$	$Poisson(\lambda)$	$e^{\mathbf{X}\Theta}$ (exponencial)
regresión logística	$\{0, 1\}$	$Bernoulli(p)$	$\frac{1}{1+e^{-\mathbf{X}\Theta}}$ (sigmoide)

¿Cómo se estiman los parámetros  $\Theta$  del modelo? Los algoritmos y heurísticas de optimización son muchos. Una vez más, aquí se describirá la forma más propia y general de asignar valores a variables no libres de estos o cualesquiera otros modelos estadísticos paramétricos: la estimación por máxima probabilidad *a posteriori* (Maximum A Posteriori: MAP), y el subcaso de estimación por máxima verosimilitud (Maximum Likelihood Estimation:

<sup>16</sup>Inventado por Nelder, Wedderburn (1972).

<sup>17</sup>La idea de ajustar modelos lineales de regresión (y regresión lineal múltiple) puede ser trazada hasta los trabajos en astronomía matemática de Legendre (1805) y Gauss (1823). Más tarde serían retomados y elaborados en un contexto genético y estadístico por Yule (1897), Pearson (1903) y Fisher (1922).

MLE).

La estimación está constreñida a trabajar óptimamente con datos limitados. Imagine que conoce la distribución de probabilidad de observar ciertos valores de los parámetros del modelo, para todas las posibles muestras de datos que de él emanaron ( $P(\Theta|Y)$ ). Entonces lo mejor que se podría hacer es tomar los parámetros más probables. No obstante, eso es exactamente lo que se ignora. No obstante, se puede recurrir al teorema de Bayes para descomponer la distribución condicional:

$$\overbrace{P(\Theta|Y)}^{posterior} = \frac{\overbrace{P(Y|\Theta)}^{verosimilitud} \overbrace{P(\Theta)}^{a\ priori\ (sesgo)}}{\underbrace{P(Y)}_{evidencia}} \quad (2.8)$$

$P(Y)$  es uniforme si la respuesta realmente es lineal y si se muestrea uniformemente en  $\mathbf{X}$ , por lo que su efecto sólo es el de normalizar el numerador, y se ignora. A menos que se quiera incorporar un sesgo con conocimiento previo acerca de  $P(\Theta)$  — razón de ser de la tradición estadística bayesiana — se asume que su distribución es uniforme y también se ignora. Lo único que resta es modelar la verosimilitud y maximizarla.

A continuación se expone el método de MLE para el caso de regresión lineal simple (una sola variable y un solo parámetro  $\theta$  en representación de la pendiente). Luego se extenderá la solución para todo modelo lineal general.

Se asumirá que los ensayos de medición son independientes (la probabilidad de observar algún valor no depende de lo observado en otros ensayos, ó  $P(y_i|y_j) = P(y_i)$ ); de manera que se pueda aplicar la regla del producto para simplificar el cálculo de probabilidades conjuntas. Expresado en notación matemática,  $P(y_i, y_j)$  en general es  $P(y_i|y_j)P(y_j)$ , pero la independencia lo reduce a  $P(y_i)P(y_j)$ :

$$P(\mathbf{y}|\theta) = P(y_1, \dots, y_n|\theta) = \prod_{i=1}^n P(y_i|\theta) \quad (2.9)$$

Supongamos que las mediciones  $y_i$  para un mismo valor de  $x$  se ciñen a un ruido gaussiano, con media en el valor verdadero según la recta de regresión. Esto ocurre sin ataduras al valor que  $\theta$  resulte tener. Supongamos además que la varianza del ruido es la misma a lo largo de la variable explicativa, condición conocida como *homocedasticidad*, que permite repetir exactamente la misma distribución de probabilidad en cada factor. Así se otorga forma concreta a la función de verosimilitud:

$$P(\mathbf{y}|\theta) = \mathcal{L}(\mathbf{y}, \theta) = \prod_{i=1}^n \mathcal{N}(\mu = x_i\theta, \sigma^2) \quad (2.10)$$

Por conveniencia analítica y numérica,<sup>18</sup> se acostumbra trabajar con el logaritmo natural de  $\mathcal{L}$ . Ya que los logaritmos son funciones monotónicas, se garantiza que ambas optimizaciones produzcan resultados idénticos:

$$\begin{aligned} \ln(\mathcal{L}(\mathbf{y}, \theta)) &= \ln\left(\prod_{i=1}^n \mathcal{N}(\mu, \sigma^2)\right) = \sum_{i=1}^n \ln\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2}\right] \\ &= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \underbrace{x_i\theta}_{\mu}\right)^2 \end{aligned} \quad (2.11)$$

Encontrar el parámetro que maximiza la log-verosimilitud equivale a encontrar dónde la derivada de la log-verosimilitud respecto al parámetro es cero.  $\ln(\mathcal{L})$  como función de  $\theta$  claramente es una función parabólica cóncava; por lo que el máximo es único, ergo también hay un solo modelo lineal óptimo. Aunque no se demostrará aquí, de hecho el máximo se obtiene sin percances para cualquier distribución de la familia exponencial (Nelder, Wedderburn 1972); aunque no siempre esté disponible en forma de expresión analítica, teniendo que echar mano de optimización numérica. Esta propiedad de todos los modelos lineales generalizados es atractiva computacionalmente. Nótese también que desde el punto de vista de la derivada, que eliminará el primer término por ser constante, se ha obtenido la misma función de costo que si se hubiera elegido arbitrariamente el error cuadrático medio (MSE) de los residuos, como se hace con estimación ordinaria por mínimos cuadrados (OLS). A saber, se está hablando de la función objetivo  $\frac{1}{n} \sum_{i=1}^n (y_i - x_i\theta)^2$ .

De vuelta con la derivada y sustituyendo  $\ln(\mathcal{L})$  por el resultado de la ecuación (2.11):

$$\begin{aligned} \hat{\theta} &= \arg_{\theta} \max \ln(\mathcal{L}) = \left\{ \theta : 0 = \frac{\partial \ln(\mathcal{L})}{\partial \theta} \right\} \\ &= \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{x} \cdot \mathbf{x}} = \frac{\text{Cov}(x, y)}{\text{Var}(x)} \end{aligned} \quad (2.12)$$

---

<sup>18</sup>Retrospectivamente al uso de MLE, los fundamentos de la teoría de la información ofrecen fuertes argumentos para trabajar con logaritmos de probabilidades. No sorprendentemente, existe una equivalencia entre MLE y estimación por minimización de información. Ortogonalmente, la mayoría de los lenguajes de programación ofrecen números de punto flotante para implementar métodos numéricos. Este tipo de dato representa números reales con precisión limitada, y el error acumulado de multiplicarlos resulta peor que el de sumarlos.

La expresión anterior es conocida como *ecuación normal*, y permite ajustar óptimamente la pendiente del modelo en un solo paso. Podría seguirse un esquema de demostración similar para obtener que el estimador por máxima verosimilitud de la varianza del ruido gaussiano es, no sorprendentemente, la conocida fórmula de la varianza:

$$\begin{aligned}\hat{\sigma}^2 &= \arg_{\sigma^2} \max \ln(\mathcal{L}) = \left\{ \sigma^2 : 0 = \frac{\partial \ln(\mathcal{L})}{\partial \sigma^2} \right\} \\ &= \frac{1}{n} \sum_{i=1}^n (y - \mu)^2 = \text{Var}(y)\end{aligned}\tag{2.13}$$

Teniendo la pendiente  $\theta$  y la media  $\mu_y$  observada empíricamente para cierto valor fijo de  $x$ , obtener el parámetro de ordenada al origen se vuelve trivial. Éste parámetro también podría obtenerse durante la mismo paso de MLE, añadiendo otra entrada al vector  $\boldsymbol{\theta}$  en un modelo lineal general, haciendo que su respectiva columna en la matriz de diseño sea  $(1 \dots 1)^T$ . Se puede probar usando los mismos razonamientos que el modelo lineal general o de regresión lineal múltiple tiene una ecuación normal análoga a la ecuación (2.12) de la regresión lineal simple. Para ser precisos:

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}\tag{2.14}$$

O más generalmente, para ajustar todos los modelos simultáneamente en una operación matricial (por ejemplo, cuando se quiere ajustar el GLM de cada serie de tiempo):

$$\hat{\boldsymbol{\Theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}\tag{2.15}$$

### 2.3.2. Neuroimagen funcional como un sistema dinámico

En cambio se podrían considerar todas las señales cerebrales simultáneamente como sistema, y tratar de predecir con ellas la manipulación experimental: para cada instante, cada una de estas señales aporta una coordenada para localizar el estado actual del sistema. Entonces podemos pensar que el video completo de fMRI es equivalente a un punto que se mueve en un espacio de muchas dimensiones, tantas como voxels, y su trayectoria en éste podría ir prefiriendo zonas según cambian las variables experimentales. Si se busca predecir valores continuos en la variable dependiente entonces se sigue hablando de métodos de regresión multivariada (de hecho GLM podría llegar a usarse de esta manera poco canónica para neuroimagen funcional. El estudio de Stringer et al. (2019) es muestra de ello). Si la variable a predecir toma valores discretos entonces no se habla de regresión,

sino de clasificación. Más al respecto en breve.

Esta representación geométrica del sistema dinámico es conocida como “espacio de estados”, o “espacio fase” en física. El espacio fase contiene exactamente la misma información que la secuencia de imágenes. Todo abordaje analítico a los datos tiene un análogo en espacio fase. Por ejemplo, un análisis de conectividad funcional mediante correlación lineal es reducible a cuantificar la preferencia del sistema a moverse a lo largo de diagonales entre los ejes. Como veremos enseguida, el análisis clásico por GLM tiene cabida de la misma manera usando esta representación.

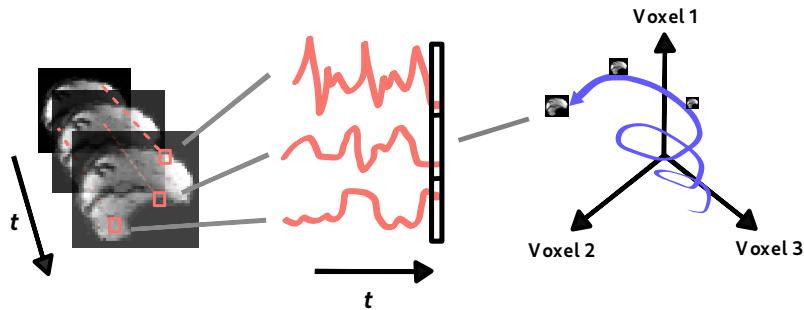


Figura 2.8: Representando la función cerebral como un sistema dinámico. Una secuencia de imágenes funcionales (izquierda) es equivalente a la trayectoria de un punto en espacio fase (derecha), con una dimensión distinta para cada señal BOLD (centro). Aquí solo se han seleccionado tres dimensiones.

### 2.3.3. Mapas de activación: análisis clásico vs MVPA

En este punto se vuelve clara la limitante del análisis por GLM: asume que el correlato neural yace en sitios individuales. Siguiendo con nuestra representación de varios voxels, podemos fácilmente concebir patrones de actividad cerebral para los cuales GLM es más que suficiente, pero también otros para los que fracasaría en distinguir los estados.

Los diversos métodos multivariantes han demostrado tener mayor sensibilidad extendiendo la identificación de interacciones complejas de actividad cerebral que no son abordables mediante modelos de regresión tradicionales (Huettel et al. 2009; Mahmoudi et al. 2012; Lewis-Peacock, Norman 2013). Agrupados bajo el nombre de “análisis de patrones multivoxel” en la literatura neurocientífica (MVPA por sus siglas en inglés), todos ellos comparten la característica de modelar la codificación del estado mental, cognitivo o conductual como una función de muchas dimensiones; que en el caso de fMRI toman la forma de voxels distantes en espacio o tiempo (Norman et al. 2006; Huettel et al. 2009; Mahmoudi et al. 2012; Lewis-Peacock, Norman 2013)

Considere el siguiente ejemplo en el que el patrón de activación (ej., la señales de actividad en distintas áreas del cerebro) bajo dos condiciones o macroestados experimentales, llamados *A* y *B*, ha sido construido según la siguiente relación (reducida a dos voxels por

simplicidad):

$$\begin{cases} A : voxel_1^2 + voxel_2^2 > r^2 \\ B : voxel_1^2 + voxel_2^2 < r^2 \end{cases} \quad (2.16)$$

Esto no es más que un círculo en el espacio fase. Si restringimos los valores a un intervalo finito y constante centrado en  $(0, 0)$ , llamado  $c$ , entonces podemos obligar a los macro-estados  $A$  y  $B$  a poseer el mismo número de puntos; si y solo si  $r$  en la expresión (2.16) cumple que:

$$\frac{c^2}{2} = \pi r^2 \quad (2.17)$$

Luego, se muestrean algunos puntos aleatoriamente de manera uniforme (40 en las figuras 2.9 y 2.10). Al graficarlos en el plano, se observa que nuestras mediciones contienen suficiente información para distinguir ambos estados cerebrales, incluso por simple inspección visual. Aunque la actividad de ningún voxel correlaciona con que el estímulo haya sido “círculo” o “triángulo”, un patrón discriminante emerge de la actividad conjunta. MVPA no es más que el uso de algoritmos de aprendizaje automático para distinguir esos patrones. En el caso del aprendizaje supervisado o clasificación, lo que se pretende es construir una frontera capaz de catalogar los datos.

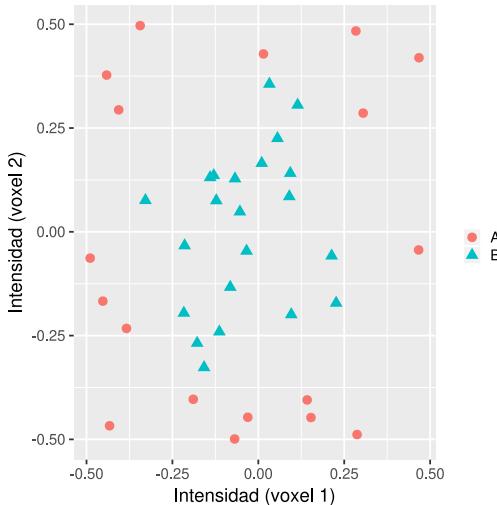


Figura 2.9: Patrón bidimensional no lineal de actividad de dos voxels, según se muestreó aleatoriamente con 18 ensayos bajo la condición experimental A más 22 bajo la B. Las mediciones de la señales que corresponden a la condición A pueden distinguirse por su excentricidad con respecto de la condición B.

Sin embargo un análisis univariado (como regresión mediante GLM) está destinado a fracasar (ver figura 2.10). Las distribuciones unidimensionales de muestra se encuentran aproximadamente centradas alrededor del mismo valor de intensidad, así que no se observa

ningún efecto intercondición. Por otro lado, las mediciones están muy dispersas y lucen ruidosas. Tal enormidad en el rango de valores para cada estado tampoco puede ser explicado por artefactos de movimiento ni por un proceso determinista no estacionario que empuje sistemáticamente los valores; ya que el proceso usado para generar los estados fue explícitamente ninguna de esas cosas.

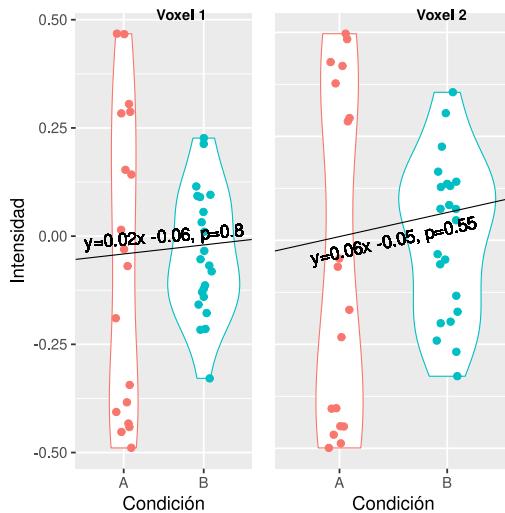


Figura 2.10: Modelos de regresión lineal simplificados y valores  $p$  de las pruebas  $t$  para las proyecciones de los datos de la figura 2.9 sobre cada eje. El efecto de las condiciones experimentales sobre voxels individuales es negligible (atestiguado por las pequeñas pendientes de los modelos), y probablemente son resultado del azar (valores  $p$ ), como se esperaba. También por construcción, una prueba de Shapiro-Wilk revela que las distribuciones de datos A no lucen gaussianas, haciendo más cuestionable la aplicación de este tipo de análisis.

Incluso un clasificador univariado no lineal cometería una gran cantidad de errores en la intersección entre ambas distribuciones.<sup>19</sup> Un patrón discriminante univariado podría observarse a partir de un análisis dinámico del espectro de frecuencias bajo condiciones dinámicas especiales,<sup>20</sup> pero el enfoque más adecuado por mucho es considerar ambos voxels simultáneamente; como en la figura 2.9, ya que así es como se generaron los datos en primer lugar.

Los beneficios son claros, pero no se adquieren sin desventajas. En primer lugar el cómputo se vuelve más complejo. Ya no se analizan voxels, sino combinaciones de voxels, algunas útiles para embargar buen desempeño del modelo predictivo. Muchas otras sólo serán detrimetiales. Bastan unos cuantos cientos de señales para que el número de combinaciones rápidamente exceda el número de átomos en el universo observable. Las toscas imágenes de fMRI suelen contener decenas de miles de señales.

<sup>19</sup>Imagine que todos los puntos de un voxel son colocados sobre la recta real, y permitimos dibujar una curva continua sobre todo el plano, con tal de separar las dos condiciones.

<sup>20</sup>Si las oscilaciones de la trayectoria durante un estado fueran más rápidas, por ejemplo. O más generalmente, en la medida en que su descomposición de Fourier favorezca frecuencias exclusivas.

### **2.3.4. MVPA en neuroimagen funcional afectiva**

Dada esta limitante en los métodos clásicos de análisis de patrones diferenciales de activación en fMRI y PET, el campo de neuroimagen funcional orientada a tareas comienza a recurrir a análisis verdaderamente multivariados para dilucidar las preguntas de interés. Al considerarlas holísticamente, MVPA es capaz de extraer igual o mayor información de las burdas imágenes funcionales en la mayoría de los casos (ver Jimura, Poldrack 2012 para contraejemplos), a expensas de complicar potencialmente el modelo y su cómputo. La adopción de este enfoque en estudios con fMRI ha llevado a descubrimientos novedosos en casi todas las ramales de la neurociencia cognitiva (por ejemplo Haynes, Rees 2005; Polyn et al. 2005; Soon et al. 2008; Rissman, Wagner 2012; véanse Huettel et al. 2009; Tong, Pratte 2012 para más).

Según Lewis-Peacock, Norman (2013), el término MVPA es usado para referirse a dos estrategias distintas de análisis de patrones multivoxel; ambas provenientes del campo de estudio en el que se intersecan la estadística inferencial multivariada y la inteligencia artificial, un área conocida como “aprendizaje de máquinas” o “aprendizaje automático” (“machine learning” en inglés):

- Clasificación automática empleando algoritmos de aprendizaje supervisado.
- Cuantificación de similitud entre patrones, según alguna métrica de distancia.

Mientras que los primeros decodifican estados preestablecidos a partir de la actividad registrada, la segunda línea de análisis se asemeja a la teoría de los modelos de aprendizaje automático no supervisado; y puede entre otras cosas usarse para discernir si dos zonas del sistema nervioso codifican información similar (Lewis-Peacock, Norman 2013).

Para 2012 ya se habían realizado un puñado de estudios de decodificación emocional con análisis multivariado, y su potencial era reconocido por la comunidad en trabajos revisorios (como Hamann 2012; o Schirmer, Adolphs 2017 en lo que respecta a percepción emocional; ver Kragel, LaBar 2014; y Kragel, LaBar 2016 para revisiones abocadas al uso de MVPA en materia de emociones). También es necesario recalcar que aunque muchas de las aplicaciones de MVPA en neurociencia afectiva han usado percepción de estímulos emocionalmente cargados, ésta no es exactamente análoga a la experiencia en primera persona de una emoción (notable también a partir de actividad cerebral (Wager et al. 2008)). La percepción de emociones en estímulos sociales además puede involucrar procesos de evaluación, atribución de estados mentales e intenciones; a la vez que puede estar exenta de emoción *per se* (debido a la modulación empática y sus trastornos) (Peelen et al. 2010). Con todo y eso, ambos procesos parecen recurrir a los mismos circuitos de evaluación de

carga emocional. En lo subsecuente se hará la distinción, poniendo énfasis en los trabajos relativos a *clasificación* de estados de *percepción* emocional.

La tabla 2.2 enlista todos los estudios encontrados para clasificación de percepción emocional. Éstos pueden ser tamizados según la modalidad de los estímulos, las emociones probadas, si la búsqueda se concentra en regiones de interés o si se hizo algo más parecido a un análisis de cerebro completo. También, por el algoritmo de clasificación, y obviamente por sus resultados. La forma más sucinta de expresar el resultado es comparando la capacidad de clasificar correctamente los patrones cerebrales (exactitud) respecto de un desempeño por azar. Observamos que si bien no son perfectos, siempre se ha reportado evidencia de estar por encima del azar.

Pessoa, Padmala (2007) demostraron la factibilidad de usar MVPA para predecir la respuesta conductual en una tarea de fMRI sobre percepción de caras asustadas, presentadas casi subliminalmente, en contraste con otros tipos de caras. Se basó en un algoritmo de entrenamiento SVM sin kernel, cuyo desempeño aumentaba conforme las variables independientes en el modelo eran adicionadas con nuevas regiones de interés (ROIs por sus siglas en inglés); hipotetizando que los varios aspectos de la actividad emocional (discriminación del estímulo, producción de respuestas, etc.) contienen información útil para recuperar el estado perceptual. Ethofer et al. (2009) utilizaron estímulos auditivos con carga paralingüística emocional. Su modelo SVM fue capaz de clasificar entre cuatro emociones básicas y un estado aemocional por encima del azar, mientras que un análisis univariado no encontró diferencias en las mismas señales BOLD, provenientes de corteza auditiva (ver Kotz et al. (2012) para un estudio similar no restringido a corteza auditiva). Said et al. (2010) utilizaron clasificación multivariada mediante el método de regresión logística para demostrar que la actividad de ciertas regiones del surco temporal superior (STS, previamente vinculadas a detección de expresiones faciales) predice la categoría emocional de los rostros. Los mismos datos bajo análisis univariado fueron incapaces de repetir la hazaña. Peelen et al. (2010) probaron estímulos de cinco categorías emocionales en modalidades distintas (audición, visión de gesticulaciones y de lenguaje corporal), y realizaron una búsqueda de regiones codificantes en todo el encéfalo mediante el método de “searchlight”. Sorprendentemente, el uso de corteza prefrontal medial (mPFC), o bien del STS, resultó capaz de hacer la distinción entre emociones, independientemente de la modalidad sensorial e intensidad del estímulo. En cambio, la aplicación de GLM sólo replicó este hallazgo supramodal en el caso del miedo. Wegrzyn et al. (2015) exploraron varias ROIs<sup>21</sup> asociadas a procesamiento de rostros para comparar la magnitud en la que contienen información emocional. Todas resultaron aptas, especialmente el STS y la bien conocida FFA (especialmente la derecha, convergiendo con el conocimiento previo).

---

<sup>21</sup>Region of Interest: región de interés.

El uso de MVPA en predicción de percepción emocional también ha alcanzado a otros modelos de las emociones y a otras especies. Skerry, Saxe (2015) utilizaron SVM para probar que las emociones percibidas, entendidas como “appraisals” o evaluaciones sociales finas y dependientes del contexto, pueden decodificarse por encima del azar a partir de redes cerebrales asociadas a la facultad de teoría de la mente. Hernández-Pérez et al. (2018) (prepublicación) entrenaron perros para permanecer quietos y despiertos en un resonador magnético mientras observaban rostros humanos felices o inexpresivos. En un segundo experimento emplearon más emociones y SVM para proveer evidencia de que los correlatos encontrados en la corteza temporal derecha del perro contienen información selectiva a la felicidad de los humanos.

En lo que respecta a experimentación en primera persona de emociones, los métodos de clasificación multivariada han sido capaces de caracterizar con precisión los dispersos correlatos espaciales de la valencia hedonística y el nivel de excitación (arousal) postulados por hipótesis dimensionales (Rolls et al. 2009; Baucom et al. 2012; Chikazoe et al. 2014; Shinkareva et al. 2014; Chang et al. 2015). Además hay evidencia de cierta generalidad intersujeto, pues existen patrones compartidos que permiten su clasificación a niveles de exactitud no triviales. De la misma forma se han encontrado resultados prometedores categorizando experiencias de emociones discretas (así se quieran conceptualizar como básicas o constructos), en comparación con los resultados del cúmulo de literatura con análisis univariado (Sitaram et al. 2011; Kassam et al. 2013; Saarimäki et al. 2015; Kragel, LaBar 2015). Está por verse si el uso de MVPA será suficiente para encontrar una hipótesis vencedora.

Cuadro 2.2: Muestra representativa de estudios experimentales sobre los correlatos neurales de la percepción de emociones que han empleado MVPA.

Estudio	Modalidad de estímulo	Emoción de estímulo	ROI	Algoritmo	Desempeño
Pessoa, Padmala (2007)	visual	miedo	varias	SVM	78 % > 50 %
Ethofer et al. (2009)	auditiva	alegría, enojo, tristeza, alivio	corteza auditiva	SVM	33 % > 20 %
Said et al. (2010)	visual	alegría, enojo, tristeza, asco, sorpresa, miedo	surco temporal superior	regresión logística	22 % > 14 %
Peelen et al. (2010)	auditiva, visual (gesticulaciones, lenguaje corporal)	alegría, enojo, tristeza, asco, miedo	encéfalo (searchlight)	no especificado	no especificado
Kotz et al. (2012)	auditiva	alegría, enojo, tristeza, sorpresa	encéfalo (searchlight)	SVM	28 % > 20 %
Skerry, Saxe (2015)	lectura, visual (caras)	21 distinciones finas contexto-dependientes (appraisals)	red de teoría de la mente, encéfalo (searchlight)	SVM	~8 % > 5 % (ROIs)
Wegrzyn et al. (2015)	visual (caras)	alegría, enojo, miedo	varias	MGPC	~32 % > 25 %
Hernández-Pérez et al. (2018)	visual (caras)	alegría	( <i>canis familiaris</i> ) corteza temporal derecha, caudado. Encéfalo (searchlight)	SVM	~65 % > 50 % (ROIs)

# Capítulo 3

## Justificación

Al día de hoy, clínicos y neurocientíficos cognitivos siguen siendo bastante dependientes de reportes verbales y otras pistas conductuales para evaluar estados subjetivos. Pruebas neurales análogas deberían idearse y descubrirse, pero no solo por sus aplicaciones terapéuticas y éticas (Seth 2010; Tononi, Koch 2015; Michel et al. 2019). Más general y fundamentalmente, se requiere trabajo teórico y empírico extra para tender un puente explicativo entre los hechos de la subjetividad (mentales) y los modelos del mundo objetivo (como el conocimiento científico) (Chalmers 1995; Seth 2010; Oizumi et al. 2014). Resulta agravante para los estudios de la conciencia que ni siquiera se tengan bien identificados los correlatos neuroanatómicos más macroscópicos de prominentes estados fenomenológicos, como los sentimientos de las emociones.

Hay varios sentidos en los que las emociones básicas podrían considerarse algunos de los estados orgánicos más primitivos, después de sensaciones como el dolor (Simon et al. 2006; Simon et al. 2008) o un estado básico de vigilia reactiva/alertamiento (que Damasio 1999 denomina “emociones de fondo”): son universales culturales relativamente diferenciables entre sí (Ekman, Davidson 1994), y están inextricablemente conectadas a la homeostasis y estados somáticos de gran relevancia para el organismo; de modo que su correcto procesamiento en el sistema nervioso, como ya se mencionó durante la introducción de este trabajo, tiene una función preponderante en los mecanismos de la motivación, la toma de decisiones y el control ejecutivo. Desde el punto de vista de la percepción emocional, ayudar a entender los mecanismos en tercera persona para su detección y procesamiento contribuye a nuestra comprensión de la intersubjetividad. Así mismo, el reconocimiento de expresiones faciales es una forma primaria de comunicación no verbal que involucra la función de vastos y distribuidos circuitos neuronales, los cuales se encuentran bajo investigación activa.

Pese al éxito inicial identificando y describiendo los sustratos moleculares y celulares de algunos fenómenos psicológicos y afecciones psiquiátricas; los métodos de la neurociencia a escalas pequeñas y medianas, así como el llano *localizacionismo*, fracasan frente a cantidad de preguntas de interés. Muchos fenómenos psicológicos más bien emergen (o así se piensa) de la función coordinada de componentes anatómicamente distribuidos (Gazzaniga, Ivry 2013; Oizumi et al. 2014); y la percepción de expresiones faciales no es la excepción

(Kragel, LaBar 2014). Las restricciones evolutivas imponen límites en el número de especializaciones nerviosas que podrían corresponder unívocamente con el desempeño de una función, de manera que ciertos fenómenos cognitivos deben corresponder al reclutamiento diferencial y modulación temporal de recursos más primitivos en el dominio fisiológico (Gazzaniga, Ivry 2013). Más aún, las nuevas corrientes en estudios de la cognición hacen hincapié en considerar holísticamente no solo la actividad del sistema nervioso central como si se tratara de un sistema confinado; sino la corporización de la cognición como un mecanismo evolucionado para la acción y la predicción activa (Friston 2010), e incluso considerando las vivencias con el medio (ontogenia y filogenia) y cualquier restricción que éste llegue a imponer sobre los aparatos cognitivos (Newen et al. 2018). Naturalmente, dichos enfoques han sido bienvenidos por algunos estudiosos de la emoción de manera independiente al enfoque corporizado pionero de Damasio (por ejemplo Harrison et al. 2010; Seth 2013; Slaby 2014).

El actual debate alrededor de los mecanismos cognitivos de la percepción de expresiones faciales se beneficiará de evidencia adicional en materia de disociación anatomo-funcional; y lo que es más apremiante: el uso de métodos de análisis más sensibles para derivar la evidencia necesaria.

Paradójicamente, los estudios disponibles con MVPA han limitado sistemáticamente las bondades matemáticas de estos métodos; cuando circunvienen analizar datos de gran dimensionalidad espacial y cuando, pese a haber realizado un análisis de cerebro completo, pierden la oportunidad de realizar pruebas estadísticas rigurosas para derivar mapas anatómicos de actividad cerebral relacionada a la tarea. Ambas características son parte integral del análisis clásico, que los proponentes de MVPA no escatiman calificar como complementario, si no es que obsoleto. La gran mayoría de los estudios con MVPA más bien han enfatizado su capacidad para hacer decodificación neural y predicción a partir de un subconjunto de voxels o regiones de interés, seleccionadas *a priori* o heurísticamente, así se desperdice la oportunidad de encontrar un mejor patrón haciendo una búsqueda guiada por los datos, más extensiva y distribuida (como son los correlatos cerebrales conocidos de las emociones).

Otras causas para estas flaquezas en la literatura de MVPA son más bien limitantes computacionales y metodológicas: la adición de variables puede ser detrimental para los resultados cuando sólo algunas de ellas contribuyen información a diferencia de ruido, además de complicar el cómputo. Por lo tanto puede ser legítimo a los objetivos de un estudio buscar reducir la dimensionalidad (también conocido como “selección de características”). Tanto la reducción de dimensiones como la selección de algoritmo de aprendizaje de patrones puede que impactare también en la factibilidad de producir mapas anatómicos de actividad, ya que muchos de ellos descartan la correspondencia entre sus propias va-

riables y la topología de los voxels originales. Con todo y eso, éstas son áreas de gran oportunidad para los estudios de discriminación de emociones. De hecho, el algoritmo más recurrido (SVM) *sí tiene interpretación espacial* y ha sido subutilizada. A continuación se revisarán los mejores acercamientos hasta ahora, seleccionando de la lista mencionada en el capítulo de *Antecedentes*.

En materia de percepción emocional, Kotz et al. (2012) hicieron pruebas estadísticas a nivel de grupo, produciendo mapas anatómicos de actividad relevante a la discriminación de expresión emocional en estímulos sonoros. Sin embargo, los datos de entrada al análisis grupal provienen de una búsqueda tipo “searchlight”, que se restringe a analizar voxels contiguos dentro de una esfera de pocos milímetros de radio. Skerry, Saxe (2015) también realizaron un análisis global mediante searchlight y análisis estadístico a nivel grupal, como justificación de sus ROIs y evidencia complementaria. Hernández-Pérez et al. (2018) limitaron la elaboración de mapas anatómicos a un espacio de búsqueda con el 1% de voxels más sobresalientes según sus puntajes  $F$  en una prueba ANOVA, y no se intentó un análisis grupal, probablemente porque el poco traslape espacial tras los ANOVAs por sujeto lo volvería fútil. Los puntajes  $F$  son una razón entre la varianza media intergrupal (Mean Square Among groups) y la varianza media intragrupal (Mean Square Within groups), como define la ecuación (3.1). Asume entre otras cosas que los voxels más relevantes para lograr la decodificación son los que, vistos como distribuciones de probabilidad gaussianas, se separan más del resto. Esto merma la decisión de usar MVPA en primer lugar, que suele ofrecer mayor generalidad.

$$F = \frac{MSA}{MSW} = \frac{\frac{1}{G-1} \sum_{g=1}^G n_g (\bar{x}_{g\cdot} - \bar{x}_{..})^2}{\frac{1}{N-G} \sum_{g=1}^G \sum_{i=1}^{n_g} (x_{gi} - \bar{x}_{g\cdot})^2} \quad (3.1)$$

Donde  $G$  es el número de grupos,  $N$  el número neto de datos y  $n_g$  el número de datos para el grupo  $g$ .

En lo que respecta a experimentación de emociones en primera persona, Saarimäki et al. (2015) también filtraron sus voxels mediante puntajes  $F$ , luego derivaron mapas anatómicos para cada categoría emocional a partir de la actividad y los pesos de la red neuronal artificial entrenada para hacer la clasificación. Esta red únicamente contenía capa de entrada (con una neurona por voxel) y capa de salida (una neurona por categoría), por lo que su entrenamiento es equivalente a optimizar las combinaciones lineales de señales que “activan” una categoría; no muy distinto de lo que logra el algoritmo SVM lineal. Extrañamente, se reportó el producto del peso de la sinapsis artificial y el valor del voxel; asumiendo sin garantías que un voxel “silencioso” durante determinado estado emocional no contribuye a manufacturar dicho estado, aunque la red neuronal le hubiese dado mucho peso. Quizás el reporte más completo hasta ahora sea el de Kragel, LaBar

(2015): un estudio multimodal de inducción de emociones que además comparó el modelo de emociones básicas vis-a-vis un modelo dimensional. Utilizó un algoritmo de análisis discriminante por mínimos cuadrados parciales (PLS-DA) entrenado con los datos de todos los sujetos, de cuyos parámetros se derivaron (y validaron estadísticamente mediante técnicas de remuestreo) patrones anatómicos para cada emoción.

Muchos de estos estudios podrían haber sido menos laxos en sus pruebas estadísticas a nivel grupal; por ejemplo, a causa del método de corrección de múltiples comparaciones, umbrales de significancia y uso de distribuciones paramétricas a priori que podrían no reflejar los verdaderos niveles de clasificación por azar. Así mismo, aunque algún tipo u otro de validación cruzada contra ejemplos fuera de la muestra de entrenamiento es estrictamente observada en la literatura; aquellos que recurrieron al esquema “leave-one-out”, o con “folds” que no corresponden a las adquisiciones de datos, no reportaron haber mitigado la autocorrelación de las señales. Es decir, si datos suficientemente temporalmente cercanos a los de entrenamiento terminaran en la partición de validación, podría ser que el buen desempeño del clasificador se debiera a similitudes en las tendencias de baja frecuencia y no a un aprendizaje de las características de interés. Esto es cierto en la medida en que el funcional de correlación cruzada resulte alto.<sup>12</sup>

---

<sup>1</sup>Se puede explicar mediante la siguiente analogía: imagine que toma consecutivamente dos fotografías de un amigo, con el afán de usar la primera para pedirle a un tercero que aprenda a reconocerlo, y la segunda para poner su habilidad a prueba. Puede que acierte debido a que en la segunda foto llevaba la misma ropa, porque el fondo y los colores son similares; y no debido a que aprendió a reconocer a la persona.

<sup>2</sup>Aunque no son usados explícitamente para este propósito, el uso de filtros pasa-altas y regresión para “detrending” podrían estar mitigando el problema.

# **Capítulo 4**

## **Hipótesis**

### **4.1. Preguntas de investigación**

- ¿Es posible decodificar la emoción percibida en una expresión facial a partir de actividad fisiológica cerebral; en especial, considerando simultáneamente toda la información disponible?
- ¿Es posible obtener mapas estadístico-anatómicos de actividad cerebral relacionada a tarea mediante métodos multivariantes, considerando simultáneamente toda la información disponible?

### **4.2. Hipótesis**

- Hipótesis nula: el poder predictivo de modelos de clasificación multivariada entrenados con neuroimágenes funcionales bajo percepción de dichos estímulos es *igual* al nivel de predicción por azar.
- Hipótesis de trabajo: el poder predictivo de modelos de clasificación multivariada entrenados con neuroimágenes funcionales bajo percepción de dichos estímulos es *mayor* al nivel de predicción por azar.

### **4.3. Predicciones**

Si existiera evidencia (en las señales BOLD, por ejemplo) para preferir la hipótesis de trabajo, entonces las estructuras anatómicas involucradas (ver sección de *Antecedentes*, subsección de *Bases neurales de las emociones*) y su actividad podrían obtenerse a partir de los modelos de clasificación exitosos; ya sea a nivel individual o grupal.

# **Capítulo 5**

## **Objetivos**

### **5.1. General**

- Estudiar las bases biológicas de la percepción emocional en la macroescala de redes funcionales de cerebros humanos.

### **5.2. Específicos**

- Predecir la presentación de algunas emociones básicas a partir de datos de resonancia magnética funcional.
- Derivar imágenes de los correlatos neuroanatómicos y funcionales de esta facultad cognitiva, contribuir al campo de neurociencia afectiva con dicha evidencia.
- Construir y documentar un *pipeline* válido, general, reproducible y reutilizable para el análisis de datos de fMRI orientado a tareas por bloques mediante clasificación multivariante.
- Contrastar los resultados del método multivariante contra el análisis clásico.

# Capítulo 6

## Muestra, Materiales y Métodos

Se analizaron dos bases de datos distintas en búsqueda de evidencia de una representación neural de la emoción de los estímulos. Mientras que la primera proviene de un experimento psicológico que no involucró al autor del presente proyecto, el segundo experimento fue diseñado desde cero con el propósito expreso de mitigar las deficiencias percibidas en el primero. Ambos estudios siguen un diseño transversal de un solo grupo. En lo subsecuente serán referidos como **Experimento 1** y **Experimento 2**, respectivamente. Es importante redundar en que ambos se destinaron a responder exactamente la misma cuestión, y por lo tanto serán descritos simultáneamente, resaltando sus diferencias.

### 6.1. Participantes

El Experimento 1 reunió 34 participantes neurotípicos; sin reporte de padecer o haber sido diagnosticados con algún síndrome neurológico o psiquiátrico.

El Experimento 2 reunió 17 participantes de ambos sexos con edad promedio de 24 años, quienes también negaron padecer o en algún momento haber sido diagnosticados con alguna condición neurológica o psiquiátrica, motivo por el cual tres interesados fueron excluidos. Con excepción de un participante zurdo, todos reportaron expresar fenotipo de lateralidad diestra en algún grado. Posterior a su participación, un sujeto fue excluido del análisis debido a una falla de sincronización en la adquisición de imágenes; ya que habría conducido a un desbalance en el número de muestras disponibles para entrenar los modelos de clasificación automática.

Cuadro 6.1: Características demográficas de los 16 participantes satisfactoriamente incluidos en el Experimento 2 efectivo.

<i>n</i> = 16		
<b>Edad</b> (años)	<i>̄x</i>	<i>S.D.</i>
	25	3,01
<b>Sexo</b>	Mujeres	Hombres
	8	8
<b>Escolaridad</b> (obtenida o en curso)	Licenciatura	Posgrado
	8	8

## 6.2. Adquisición de imágenes

Tanto las imágenes del primer como del segundo experimento fueron obtenidas con un resonador magnético General Electric Discovery MR750 a 3 T en su punto de máxima magnitud de campo, dentro de la Unidad de Resonancia Magnética del Instituto de Neurobiología de la Universidad Nacional Autónoma de México. Los ecos inducidos fueron registrados con una antena de 32 canales montada alrededor de la cabeza de los participantes.

En una sola sesión por participante se adquirieron 3 (Experimento 1) ó 5 (Experimento 2) secuencias eco-planares para señal BOLD, respectivamente de 260 y 185 volúmenes c/u. Cada volumen contiene 34 ó 35 rebanadas transversales respectivamente, con una matriz de 64x64 cada una, a voxels isométricos de  $4\text{mm} \cdot 4\text{mm} \cdot 4\text{mm}$  para ambos experimentos. Además se adquirió una imagen anatómica pesada a contraste T2 y otra a T1 por sujeto. La tabla 6.2 describe los parámetros de cada tipo de secuencia.

Cuadro 6.2: Parámetros de secuencia utilizados durante el protocolo de adquisición de imágenes por resonancia magnética.

Parámetro de secuencia	EPI BOLD	T2w FSE	T1w FSPGR
Orientación de rebanadas	Axial	Axial	Axial o sagital
Rebanadas	34 ó 35	35	176
FOV	$64 \times 64$	$512 \times 512$	$256 \times 256$
Tamaño de voxel	$(4\text{ mm})^3$	$,5 \times ,5 \times ,4\text{ mm}^3$	$(1\text{ mm})^3$
Flip angle	$\pi/2$	$7\pi/9$	$3\pi/45$
TR (ms)	2000	6255	8,18
TE (ms)	30	101,4	3,19
TI (ms)			450

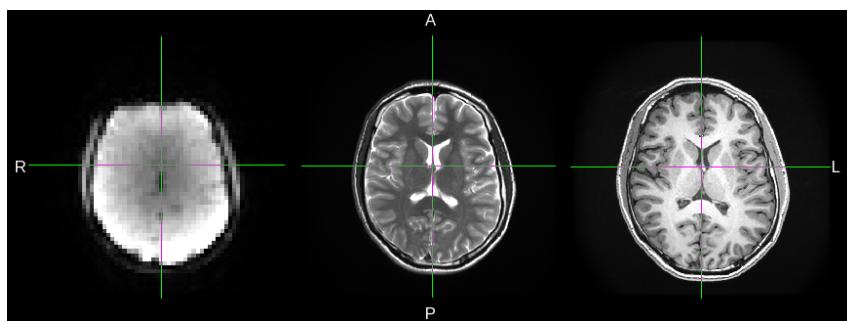


Figura 6.1: Muestra de imágenes crudas de un mismo sujeto para cada tipo de secuencia (en el mismo orden que en la tabla 6.2). Estos cortes axiales corresponden a una de las muchas rebanadas adquiridas para un mismo volumen.

### 6.3. Estímulos y tarea

Dado un rostro con mirada lateralizada, el sistema visual humano es inducido a redirigir la atención en la dirección sugerida; presuntamente porque información de relevancia ambiental y social podría ser descubierta en donde otros miran. Este cambio atencional abrupto se encuentra dentro del rango de intervalos rápidos de un movimiento sacádico. El fenómeno es conocido como “gaze cuing”, y ha sido usado para experimentar con la atención visual y su interacción con la percepción (Friesen, Kingstone 1998).

El Experimento 1 emplea una variación a la tarea de “gaze cuing” con un diseño por eventos. No obstante, para efectos de este trabajo el componente emocional de los estímulos puede considerarse como presentado en pequeños bloques de 2.7 segundos y 3 segundos, ya que la emoción es invariante dentro de cada ensayo. Cada sesión de registro comprende 126 ensayos estímulo-respuesta, donde ensayos con rostros que desvían la mirada son intercalados con otros de control (mirada directa) para evitar habituación en los sujetos y contaminación en el análisis de clasificación debido a correlación temporal. Esto se explica más a detalle en la figura 6.2: un diagrama de sistema de estados discretos de tipo markoviano, en el que se representa un par de bloques *⟨gaze cuing, control⟩*. Varias iteraciones en el diagrama estarían ocurriendo durante el registro de imágenes.

Los verdaderos estímulos provienen de la base de datos fotográfica NimStim (Tottenham et al. 2009). En la ilustración han sido caricaturizados por brevedad, ya que se utilizó más de un rostro para cada emoción (4 hombres y 4 mujeres con variantes para todas las expresiones requeridas, étnicamente europeoides, presentados de forma balanceada. Los identificadores de los modelos dentro de la base de datos son: 3, 5, 7, 8, 24, 25, 32, 36).

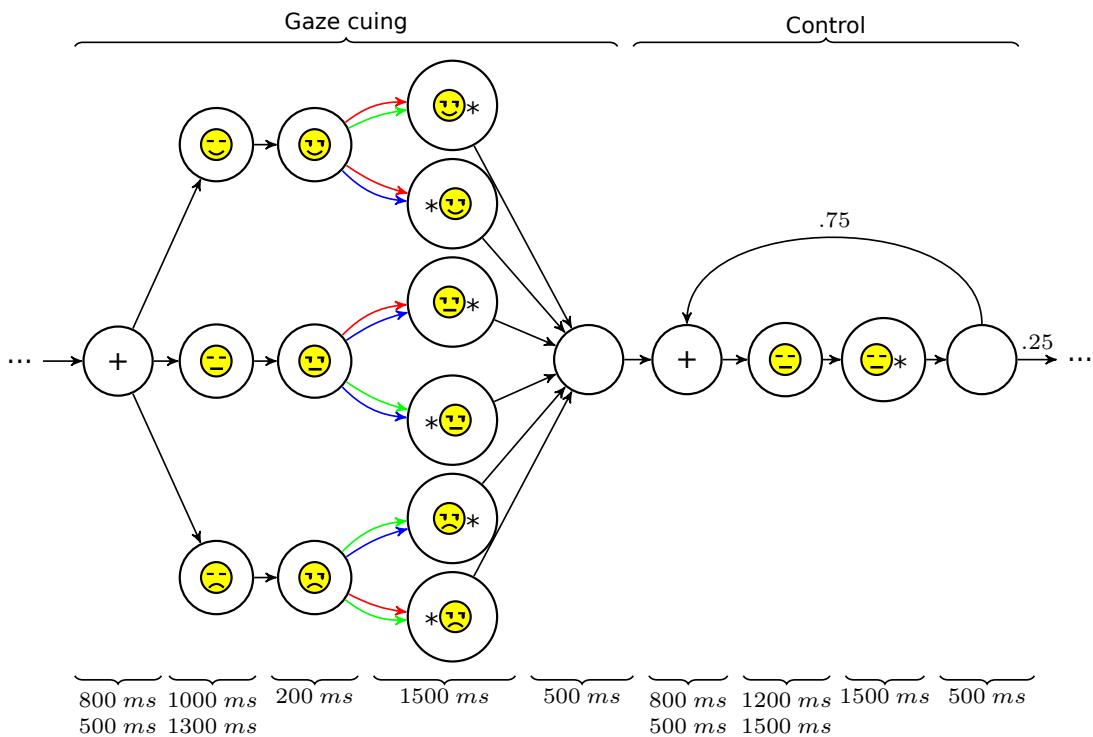


Figura 6.2: Paradigma de *gaze cuing* del Experimento 1. Los nodos representan estímulos visuales, su orden de presentación es indicado usando flechas. A menos que se indique lo contrario, la probabilidad de transición desde un nodo se distribuye equitativamente entre todas las flechas salientes (los ensayos se presentaron pseudoaleatoriamente). Se le pidió a los participantes que buscaran un objetivo (asterisco) y que presionaran el botón izquierdo o derecho, según la posición del objetivo relativo a la cara. El espectro de expresiones probadas (neutral, feliz, triste; seguidas ya fuera por un objetivo congruente o incongruente con la mirada) fue dividido y contrabalanceado en tres secuencias, con tal de mantenerlas suficientemente cortas. Las flechas coloridas indican situaciones específicas a una de las tres secuencias.

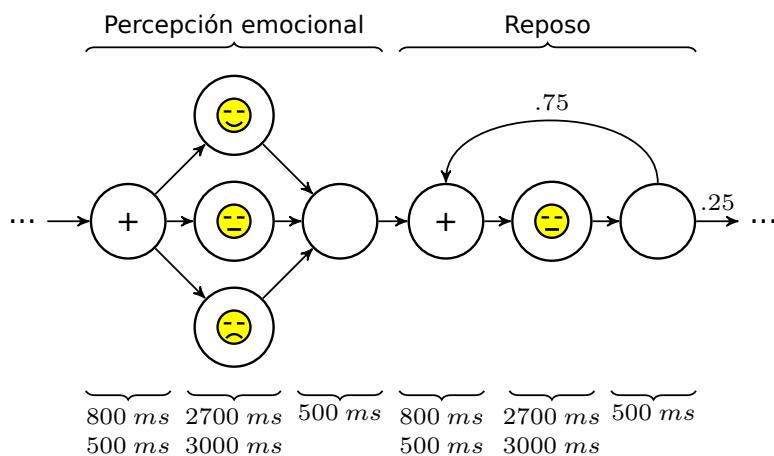


Figura 6.3: Para propósitos del etiquetado de datos de entrenamiento y prueba de clasificación mediante aprendizaje de máquinas, los estados de estimulación de la figura 6.2 pueden ser condensados en los metaestados emocionales aquí mostrados.

Para el Experimento 2 se utilizaron 5 secuencias contrabalanceadas en un diseño por bloques de 30 segundos. Las secuencias son idénticas salvo por el orden pseudoaleatorio de

presentación de los bloques. A las 3 clases de bloques del Experimento 1 (neutral, felicidad, tristeza) se añadieron otras 3: de enojo, pseudocaras revueltas como rompecabezas (“scrambled”) que conservan las mismas propiedades estadísticas de los rostros neutrales aunque ya no pueden ser reconocidas como tal, y bloques de estimulación visual mínima (“blank”). Las caras neutras o poco expresivas podrían ofrecer un mejor control contrastando entre tipos de emociones. De lo contrario sería imposible concluir a partir de un buen desempeño que el algoritmo de clasificación ha aprendido a identificar las marcas de cada emoción, en lugar de aprender a reconocer las de  $n - 1$  emociones junto con algo que no es ninguna de ellas, aunque distinto de los correlatos de la emoción restante. La elección de bloques de pseudocaras y de nula estimulación visual fue motivada por los malos resultados analizando la base de datos del Experimento 1. Mediante distinciones más triviales y de resultados bien conocidos (como simple estimulación visual) se buscó hacer una validación de todo el análisis, que al involucrar una cantidad considerable de piezas móviles y programación, ofrecía muchas oportunidades para haber cometido un errores.

A su vez, cada bloque consta de la presentación aleatoria de 10 imágenes pertenecientes a esa categoría, durante aproximadamente 3 segundos cada una y sin posibilidad de repetirse dentro del mismo bloque. Cada bloque aparecía dos veces por secuencia, con lo que se obtiene un total de 12 bloques por secuencia (360 s = 6 min). Al término de los 12, toda estimulación y petición de respuesta era interrumpida, debiendo el sujeto esperar 10 segundos antes de concluir el experimento (para dar tiempo suficiente al escáner de capturar la respuesta hemodinámica elicida por los últimos estímulos). La totalidad de una sesión para el Experimento 2 es de aproximadamente 30 minutos, 50 segundos (370 s × 5), sin considerar el tiempo de obtención de las referencias anatómicas justo antes de la tarea de fMRI.

En lo que concierne a los estímulos que conforman cada bloque, éstos fueron seleccionados de la clásica base de datos de Paul Ekman, “Pictures of Facial Affect” (Ekman 1976). En el caso de los bloques de mínima estimulación visual (“blank”) simplemente se hacía brincar aleatoriamente una cruz de fijación tenue de un cuadrante a otro dentro del espacio visual. La cruz no ocupaba más del 10 % del espacio total en cualquiera de sus direcciones. Véase la figura 6.4 para una explicación gráfica del Experimento 2.

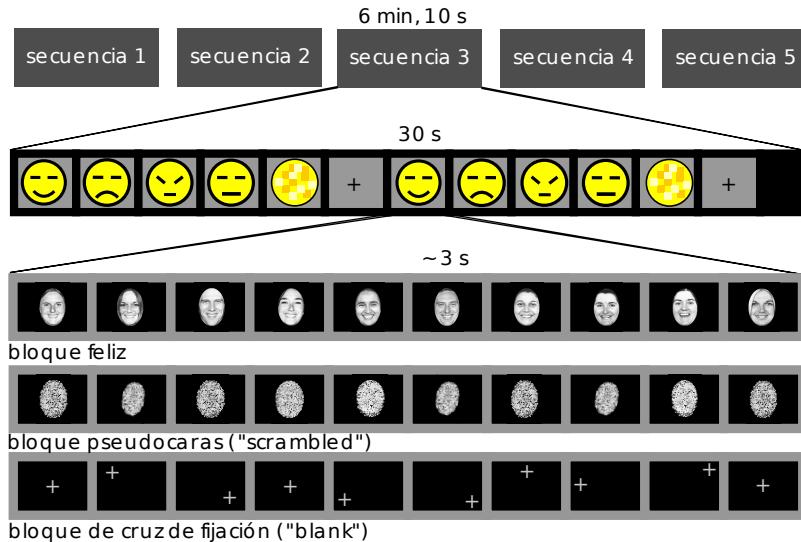


Figura 6.4: Paradigma por bloques del Experimento 2. El eje horizontal representa el paso del tiempo. Los rectángulos corresponden a unidades de estimulación (secuencias, bloques y estímulos individuales).

Se tomaron mediciones conductuales a través de la tarea como indicadores de la prestación de atención a los estímulos y así evaluar la calidad de los datos. Al inicio de cada secuencia los participantes recibieron la instrucción de indicar si cada rostro presentado pertenecía a un hombre o a una mujer. La respuesta era registrada con botones para cada dedo índice, con controles a cada mano que les permitían mantener los brazos lo más fijos posibles. En cambio, cuando la imagen claramente no correspondía a un hombre o una mujer (o persona siquiera; caso de los bloques “scrambled” y “blank”), la instrucción simplemente era reportar cada cambio en la imagen usando los índices alternadamente. De esta manera se garantizaba que la actividad motora fuera lo más homogénea posible durante todos los bloques, controlando otro posible efecto confusorio durante el análisis de clasificación de patrones pertenecientes a rostros vs bloques “scrambled”. Aunque jamás se suministró consigna explícita de catalogar o pensar en la expresión o emoción de los rostros, no es posible descartar la posibilidad de que este proceso lingüístico-conceptual apareciera espontáneamente en el tren de pensamiento de los participantes; lo cual añadiría otra posible variable de confusión.

## 6.4. Métodos de análisis

### 6.4.1. Preprocesamiento de imágenes

Los tres tipos de imágenes por resonancia magnética fueron convertidos del formato DICOM (Mustra et al. 2008) al formato NIFTI-1 (Cox et al. 2004) y estructurados en un

árbol de archivos conforme al estándar BIDS 1.4.0 (Gorgolewski et al. 2016) usando la herramienta Dcm2Bids 2.1.4 (Bedetti et al. 2019), que a su vez fue configurada para usar el convertidor dcm2niix 1.0.20170411 (Li et al. 2016) y para anonimizar los rostros de los participantes con el software pydeface 2.0.0 (Gulban et al. 2019). La sanidad de la base de datos resultante fue verificada mediante BIDS Validator 1.5.4.

Se recurrió al servicio Web de segmentación de tejidos y volumetría cerebrales Volbrain (Manjón, Coupé 2016) por sus máscaras de encéfalo (“deskulling”) y sustancia gris, y serendipitamente se aprovechó su corrección de intensidad por sesgos de campo en las imágenes T1w.

Luego las secuencias de fMRI fueron yuxtapuestas por orden temporal en una gran secuencia por sujeto, y los resultados fueron sometidos al siguiente preprocesamiento a cargo de las utilerías de FSL (Jenkinson et al. 2012): filtro de frecuencias pasa-altas ( $>50$  segundos) e interpolación para tiempo de adquisición de rebanada (Woolrich et al. 2001), corrección lineal de movimiento y corregistro lineal (12 grados de libertad, Jenkinson, Smith 2001; Jenkinson et al. 2002) con su respectiva referencia T1w y con la plantilla T1w promedio del estándar MNI-152 (Fonov et al. 2009; Fonov et al. 2011) a 1 mm de resolución.

Terminado el corregistro, se aplicó la matriz resultante correspondiente para transformar las máscaras de Volbrain al espacio de baja resolución de las imágenes de fMRI. Al sobrante de series de tiempo de origen en sustancia gris posteriormente se restan las tendencias lineales, conservando los residuos de un ajuste de regresión lineal simple por cada una de las secuencias que componen la gran serie concatenada. Finalmente y para evitar sesgar los modelos de clasificación hacia alguna dimensión, cada gran serie concatenada para cada voxel es normalizada a puntajes-z.

Para el Experimento 1, tras preprocesamiento de rutina se entrenó un clasificador SVM lineal por sujeto con los tres tipos de caras (16 muestras por clase, después de descartar volúmenes subsecuentes pertenecientes al mismo bloque, con tal de disminuir la autocorrelación de los datos). El poder predictivo intra-sujeto fue estimado usando validación cruzada “leave-one-out” (16 validaciones, usando un dato distinto a los 15 de entrenamiento como prueba en cada iteración). La exactitud de clasificación promedio luego fue comparada con una prueba no paramétrica de rangos contra la distribución de probabilidad de clasificación dada la hipótesis nula, que se estimó por el método de simulaciones de Monte-Carlo (5000 permutaciones de etiqueta en la partición de datos de entrenamiento). En lugar de asumir un retraso hemodinámico óptimo para el etiquetado de volúmenes cerebrales tras la presentación de estímulo, se midieron las series de tiempo de exactitud de clasificación por sujeto para todos los retrasos posibles (hasta 20 s) a una resolución de 200 ms.

# **Capítulo 7**

## **Resultados**

### **7.1. Experimento 1**

Para el Experimento 1, la morfología de la serie de tiempo promedio de exactitud de clasificación de 3 vías concuerda con las expectativas teóricas acerca de la función de respuesta hemodinámica; no obstante, se encontró poca o nula evidencia de decodificación exitosa, incluso tras intentar una forma simple de reducción de dimensionalidad. Sin esta reducción, la exactitud promedio de los clasificadores por sujeto alcanzó su punto máximo (28 %) tras un retraso de etiquetado cercano a los 2.5 s. Seleccionar el 1 % de voxels de sustancia gris según su puntaje-F en una prueba ANOVA empujó marginalmente el rendimiento promedio (35 %, 2.2 s y 5.2 s) por encima del nivel por azar a priori (33 %). Ver figura 7.1.

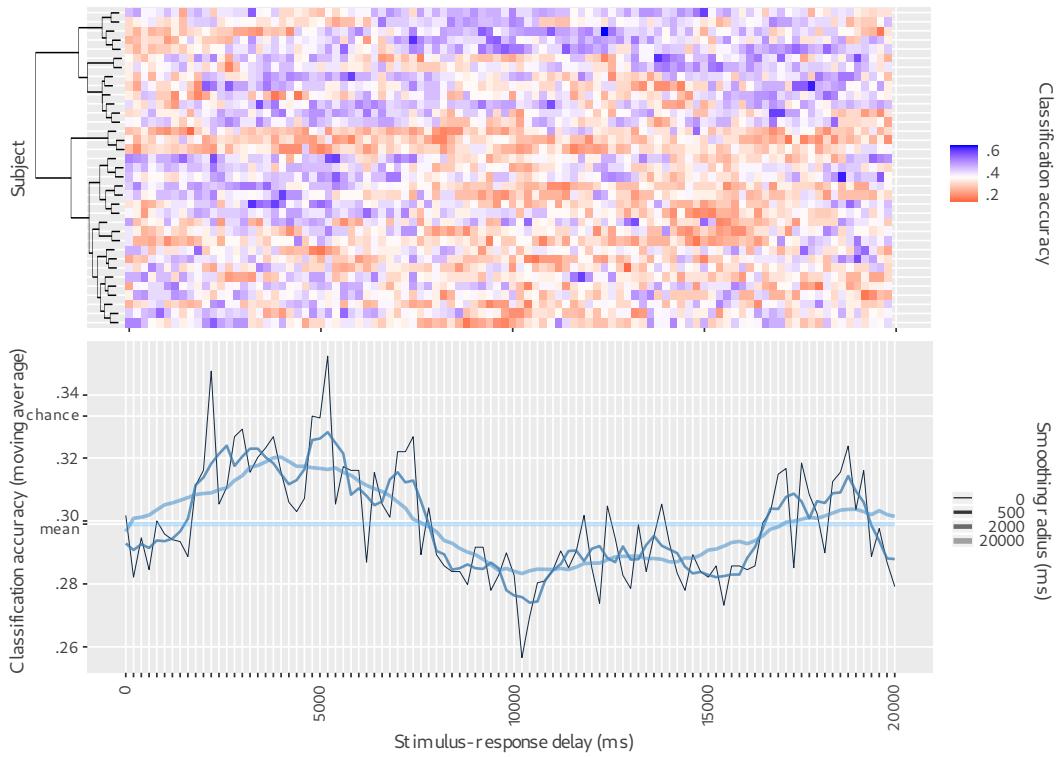


Figura 7.1: Series de tiempo de exactitud de clasificación individuales (arriba) y promedio (abajo), como función del retraso de etiquetado estímulo-respuesta. Los resultados por sujeto fueron conglomerados (“clusterizados”) mediante el algoritmo de Ward. Abajo se muestran series promedio a distintos suavizados temporales por media móvil (convolución cuadrada). Destaca una ligera tendencia a la alza pasados entre 2 y 7 segundos tras la presentación del estímulo.

Como última alternativa se permitió explorar el uso de un retraso de etiquetado distinto por sujeto. Entonces la exactitud de clasificación promedio superó el 50 % con un gran efecto estadístico ( $D$  de Cohen = 3.72), pero la probabilidad de que se debiera a una coincidencia por la exhaustividad de la búsqueda fue considerable (valor  $p$  (Bonferroni) = 0.22). Ver figura 7.2.

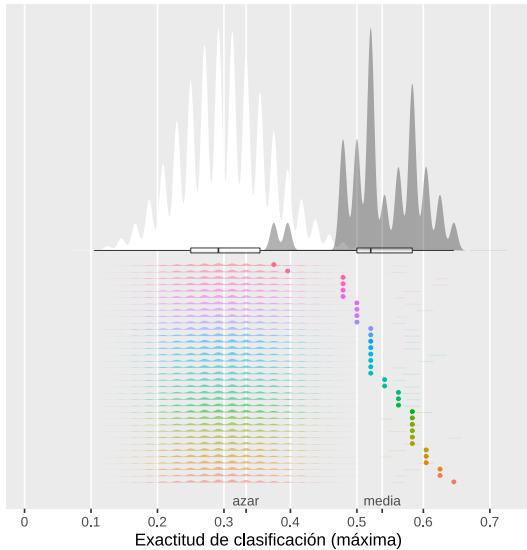


Figura 7.2: Prueba estadística por permutaciones de la exactitud de clasificación de 3 vías (feliz vs triste vs neutro), para el 1% de los voxels de materia gris más destacados según su puntaje-F. La latencia a la que fueron etiquetadas las muestras cerebrales se dejó variar por sujeto, permitiendo encontrar un gran efecto ( $D$  de Cohen = 3.72) a costa de hacer 100 búsquedas por sujeto (ver figura 7.3). La mitad inferior muestra resultados por sujeto, y la superior, el agregado grupal. La mitad izquierda corresponde a las distribuciones nulas empíricas estimadas por permutaciones, y la derecha, al valor puntual de exactitud de clasificación observada (promedio de la validación cruzada).

En resumen, la evidencia entregada por el análisis de clasificación no puede justificar una búsqueda de correlatos neuroanatómicos a nivel grupal; sin embargo se hizo una exploración cualitativa de los sujetos más ejemplares. Ésta sugiere que el mesencéfalo, porciones de los lóbulos cerebelosos anteriores, los giros fusiformes, las ínsulas, el cíngulo anterior y dorsal, la corteza orbitofrontal y el surco temporal superior podrían contener información relevante para la discriminación de rostros emocionales.

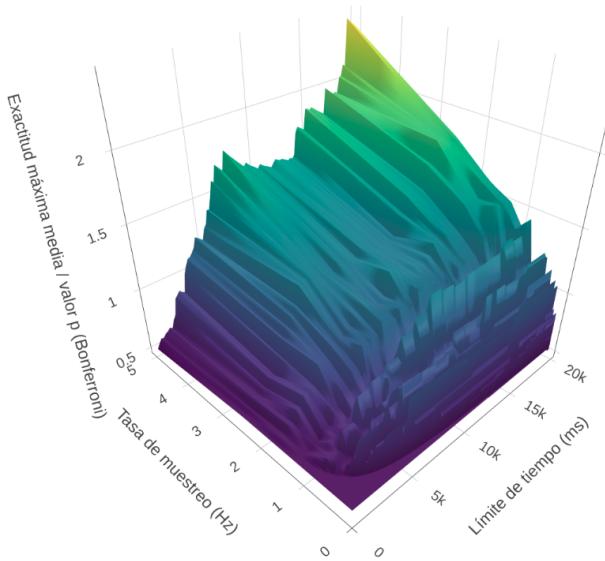


Figura 7.3: Optimización de parámetros para búsqueda de latencias óptimas individuales. ¿Cuál es la mejor relación entre exactitud de clasificación y número de búsquedas? Esta exploración de la tasa de muestreo a la que se intentan nuevos retrasos de etiquetado y el retraso máximo probado reveló que la búsqueda más exhaustiva sigue siendo más conveniente, pese a inflar los valores  $p$  por el número de comparaciones.

## 7.2. Experimento 2

### 7.2.1. Resultados conductuales

Como sucedería con cualquier experimento psicofisiológico, los datos de actividad cerebral solo son tan buenos como la tarea permita. Es importante tratar de validar las reacciones producidas por los estímulos, así como el desempeño durante la tarea, lo que puede servir a realizar un control de calidad.

Para el Experimento 2, previo al ingreso del participante al escáner se aplicó una breve prueba aleatorizada de categorización de los rostros que más tarde experimentaría durante la tarea propiamente dicha. Una prueba estadística  $\chi^2$  de Pearson para medir la fuerza de asociación entre la emoción aspirada y la interpretación subjetiva del participante reveló una minúscula probabilidad de que los participantes estuvieran respondiendo azarosamente. Aunque no todas las emociones se detectan con la misma facilidad, los porcentajes de acierto para cada una son compatibles con los reportados en validaciones hechas para otros conjuntos de fotografías de rostros emocionales.

Cuadro 7.1: Reportes externos de proporción de aciertos para las emociones bajo consideración durante el Experimento 2, comparado contra los resultados de este estudio.

Estímulos	Proporción de aciertos			
	Feliz	Enojado	Neutral	Triste
Este estudio (submuestra de Ekman 1976)	1.00	0.92	0.86	0.78
NimStim (Tottenham et al. 2009)				
(boca abierta)	0.98	0.90	0.82	0.60
(boca cerrada)	0.92	0.84	0.91	0.83
RADIATE (Conley et al. 2018)				
(boca abierta)	0.98	0.69	0.78	0.34
(boca cerrada)	0.80	0.62	0.89	0.70

De la misma manera se tomaron cálculos homólogos por sujeto (con corrección de Bonferroni para error tipo I de familia de comparaciones múltiples), pero incluso para el peor de ellos, la probabilidad de obtener tal grado de asociación respuesta-categoría según la distribución  $\chi^2$  teórica resultó menor a  $5 \cdot 10^{-8}$ .

La misma prueba grupal fue repetida separando por estímulo en lugar de agrupar por tipo de emoción. Para cada una de las imágenes, la probabilidad de obtener cierto número de aciertos en un conjunto de ensayos dicotómicos independientes (en este caso, las respuestas de todos los sujetos para dicho estímulo) está dada por la distribución binomial, y pruebas estadísticas homónimas revelaron que sólo 2 imágenes tenían una probabilidad superior a 0.05 de estar siendo clasificadas al azar (tras corrección de FWER de Holm). La prueba  $\chi^2$  de toda la matriz sigue mostrando que los bloques en general elicitaban la categoría emocional requerida, con buenos niveles de relación señal ruido para cada tipo de bloque.

### Frecuencia conjunta de categorización de estímulos

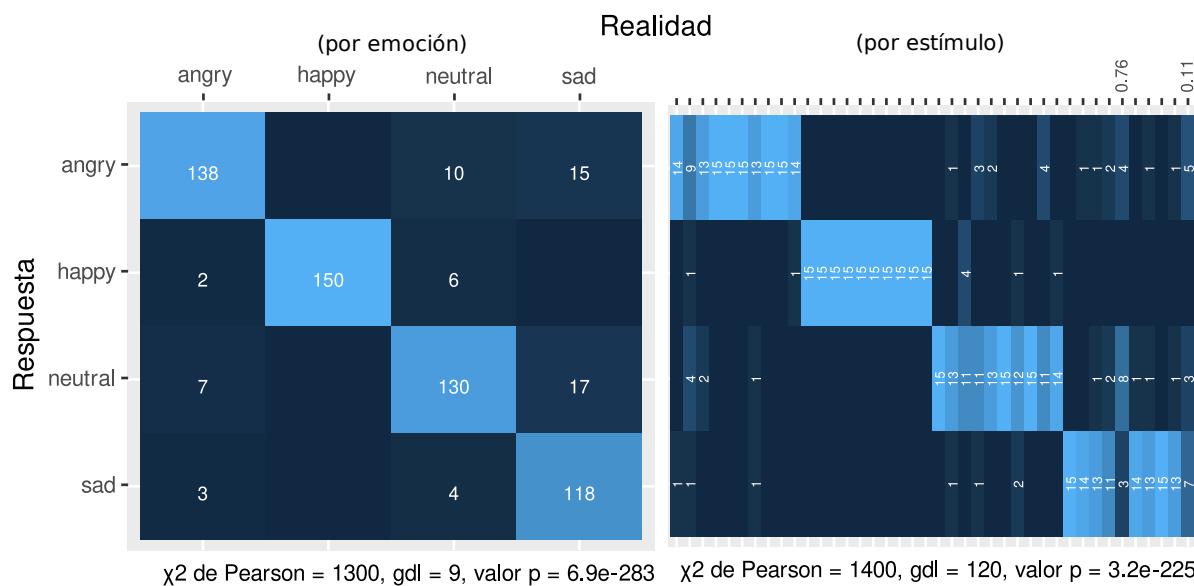


Figura 7.4: Matrices de confusión con la frecuencia conjunta de respuestas obtenidas durante la validación de los estímulos del Experimento 2. Para conjuntos de estímulos que verdaderamente cumplen su propósito se esperaría observar una fuerte diagonal, indicando que la percepción subjetiva se conforma a las categorías preasignadas. Izquierda: agrupando por emoción. Derecha: separando por estímulo para detallar la estructura de los errores. Se muestran los valores *p* corregidos (Holm) de aquellos estímulos con *p* > 0,05 según una prueba binomial de una cola bajo la hipótesis de que sólo se asigna la categoría correcta al 25 % de los ensayos de Bernoulli.

En lo que respecta a las respuestas durante la tarea, se realizaron pruebas binomiales para cuantificar la probabilidad de acierto con las botoneras, suponiendo que el nivel de acierto por azar es de 50 %. La figura 7.5 muestra el agregado de aciertos a lo largo del tiempo de los sujetos aceptados en el Experimento 2. Se distingue que los errores, en rojo, son muy pocos en proporción a los aciertos. Para el peor sujeto se encontró una probabilidad de  $6,65 \cdot 10^{-66}$  (corrección de Holm) de estar contestando al azar; y la probabilidad para el peor tipo de bloque a lo largo de la tarea, incluyendo bloques de pseudocaras y de cruz de fijación, es incluso menor.

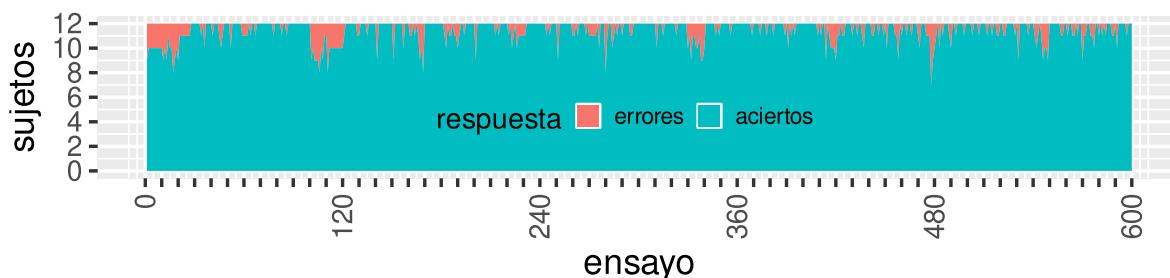


Figura 7.5: Desempeño instantáneo durante la tarea (razón entre aciertos y errores).

También se analizaron los tiempos de reacción por sujeto como medida de la prestación de atención. Cada curva de la figura 7.6 corresponde al tiempo de reacción promedio en segundos de algún sujeto, y se sombrean sus intervalo de confianza del 95 %. La recta negra superpuesta proyecta la parte relevante de un modelo de regresión de efectos mixtos lineales. El modelo de efectos mixtos es una generalización de regresión por GLM que utiliza una matriz de parámetros de efectos fijos (como en GLM), pero añade otra para modelar efectos aleatorios. Esto es especialmente apto para diseños factoriales jerárquicos; ya que el efecto de cada medición en el tiempo puede provenir de un sujeto distinto con su propia varianza personal. El ajuste usó el factor sujeto como efecto aleatorio, y el pasar de los bloques y el tipo de bloque como efectos fijos. El propósito de esto es encontrar el efecto o tendencia que tiene el transcurrir de la tarea en los tiempos de reacción. Se observa un cambio casi imperceptible en los tiempos de reacción debido al tiempo, y de hecho a la baja.

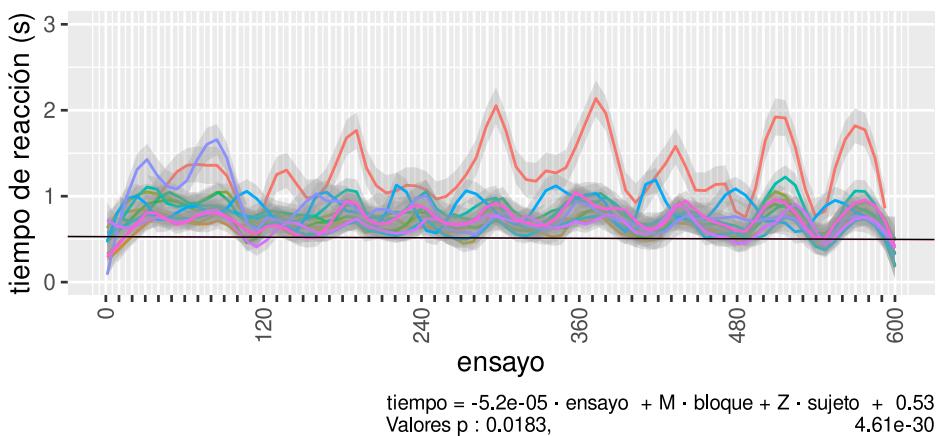


Figura 7.6: Desempeño instantáneo durante la tarea (tiempo de reacción por sujeto y ajuste lineal por modelo de efectos mixtos).

Todas estas líneas de evidencia con los datos conductuales convergen en que los sujetos comprendieron la tarea y en general prestaron atención a los estímulos. Acordemente, se optó por no descartar los datos de alguno de estos participantes, como tampoco se descartó algún tipo de bloque para el análisis de clasificación de patrones.

### 7.2.2. Validación del método

La exactitud de clasificación superó el 80 % para todos los contrastes binarios que incluían la condición “blank” en el Experimento 2. El análisis a nivel de grupo de los parámetros o pesos asignados por SVM a cada característica anatómico-funcional revela una preferencia por los voxels o parcelas situados dentro y justo alrededor de la fisura calcarina: la

corteza visual primaria (V1). Pruebas estadísticas  $z$  (sin corrección) a nivel grupal usando los pesos normalizados mediante la métrica  $L2$  únicamente superaron los umbrales recomendados para Feat (GLM+LME) en el caso del análisis reducido por parcelación, y a expensas de la sensibilidad espacial del mapa  $z$  umbralizado, que claramente sufre un deterioro en la figura 7.7. Las pruebas  $z$  para el análisis “voxel-wise” resultó simplemente infactible. El problema podría ser inherente al uso de métodos multivariantes con datos de alta dimensionalidad, como se ha venido augurando desde la sección de *Antecedentes*. Aunque es perfectamente posible explorar un número limitado de combinaciones de características muestreadas uniforme y aleatoriamente a lo largo, ancho y alto del espacio de búsqueda; la raleza (“sparsity”) entre los voxels favorecidos para cada sujeto (aunque cercanos), es tanta que no sobrevive a la gran varianza inter-sujeto. Se probaron suavizados (convolución gaussiana con FWHM de 6mm, 8mm y 10mm) de los pesos de SVM previo a las pruebas estadísticas grupales, pero una vez más, la raleza de estos parámetros merma la coincidencia espacial; haciendo de las pruebas estadísticas por voxel un pobre instrumento de detección. En su lugar, para el análisis “voxel-wise” se muestra el promedio de pesos SVM normalizados, filtrados a partir del percentil 99.

Las mismas consideraciones se repiten para el contraste “neutral vs scrambled”: evidencia fuerte de aprendizaje de los patrones característicos de cada condición según las pruebas de permutaciones por sujeto, agrupamiento coherente de los pesos asignados por cada modelo SVM hacia regiones de la red nuclear del procesamiento de caras (Occipital Face Area, Fusiform Face Area); pero con una raleza tal que la varianza por voxel es alta y el suavizado espacial en poco ayuda.

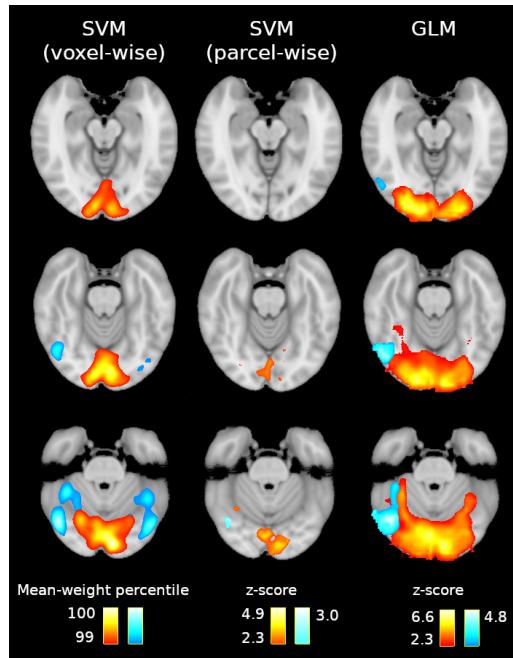


Figura 7.7: Mapas paramétricos grupales de activación para los contrastes de validación del *pipeline* de procesamiento del Experimento 2. En rojo el contraste “scrambled vs blank” (“scrambled > blank” para GLM). En azul, caras neutrales vs “scrambled” (“neutral” > “scrambled” para GLM). Destaca la sensibilidad y especificidad del análisis multivariado “voxel-wise” (sin reducción de características) filtrando los pesos de SVM normalizados a partir del percentil 99. Vista radiológica axial sobre la plantilla MNI-152.

### 7.2.3. Percepción emocional

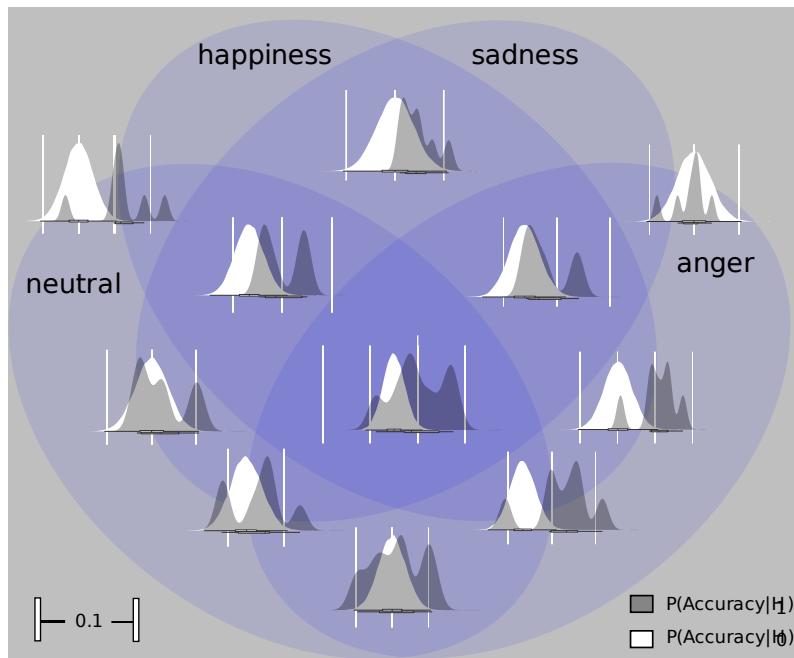


Figura 7.8: asdf.

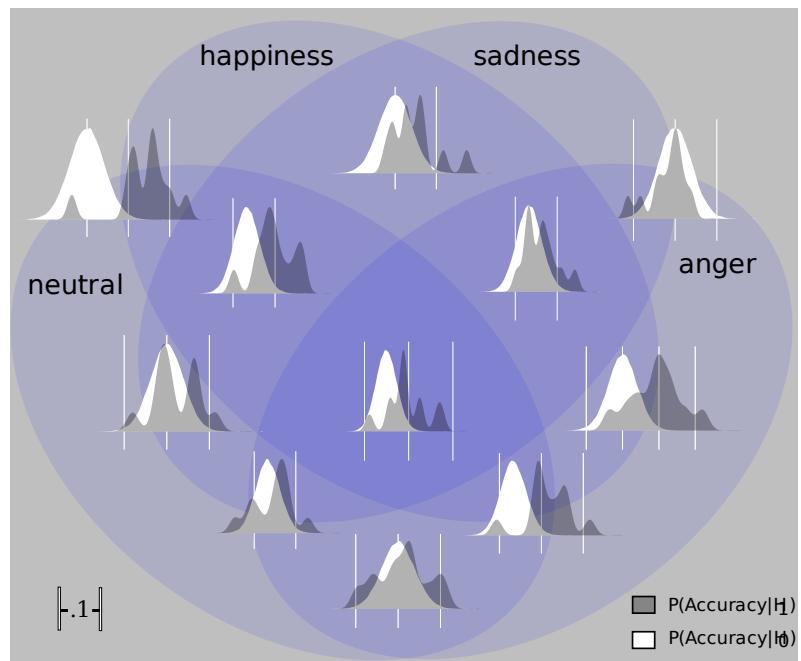


Figura 7.9: asdf.

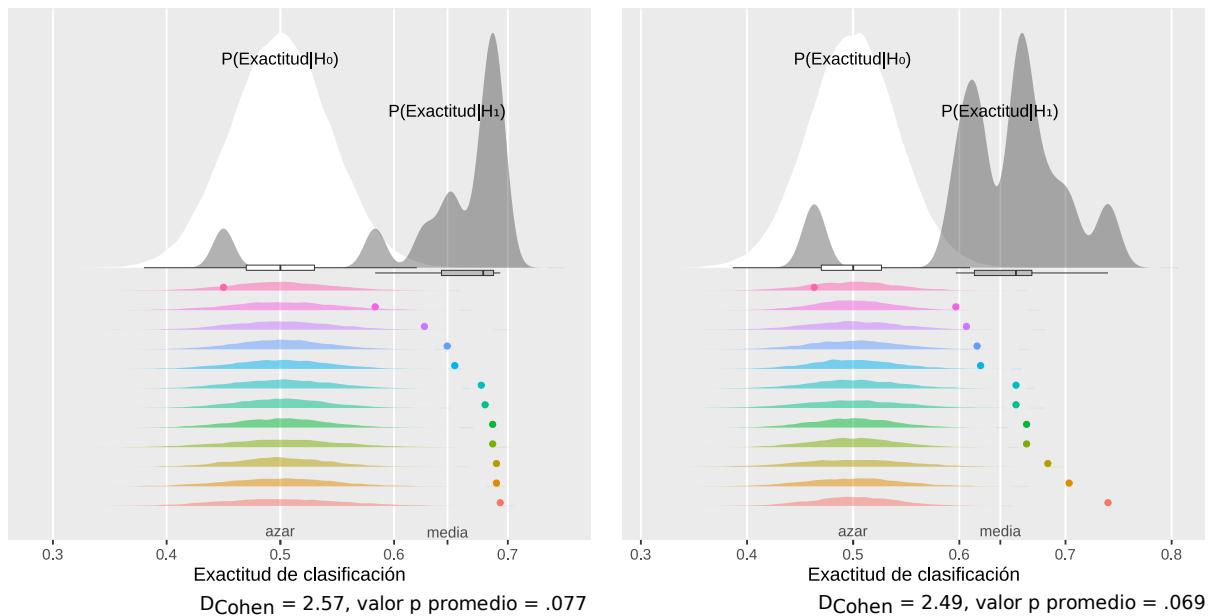


Figura 7.10: asdf.

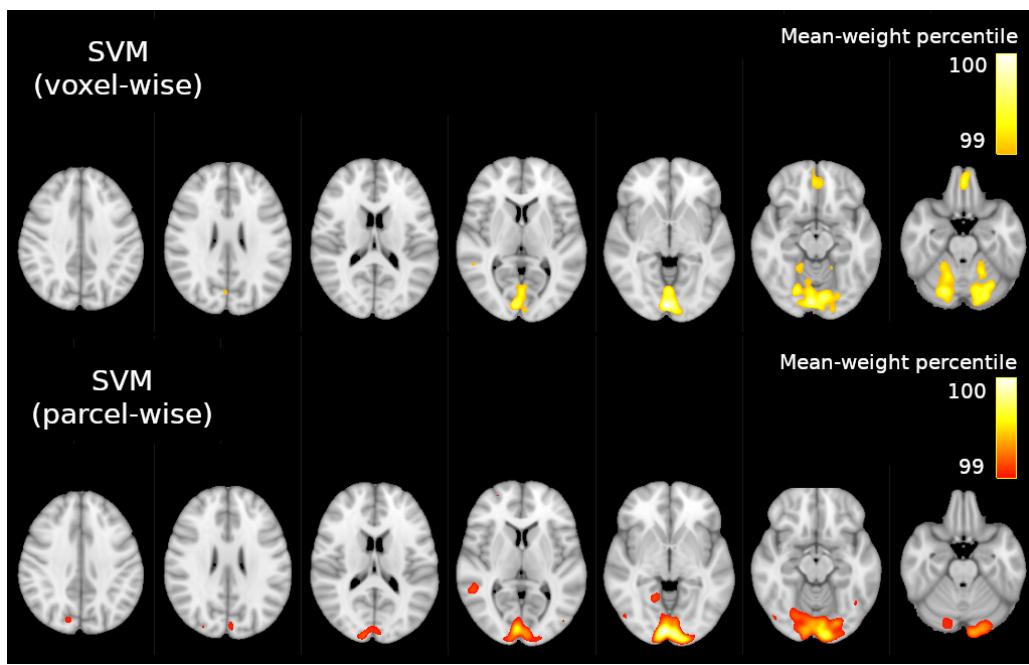


Figura 7.11: asdf.

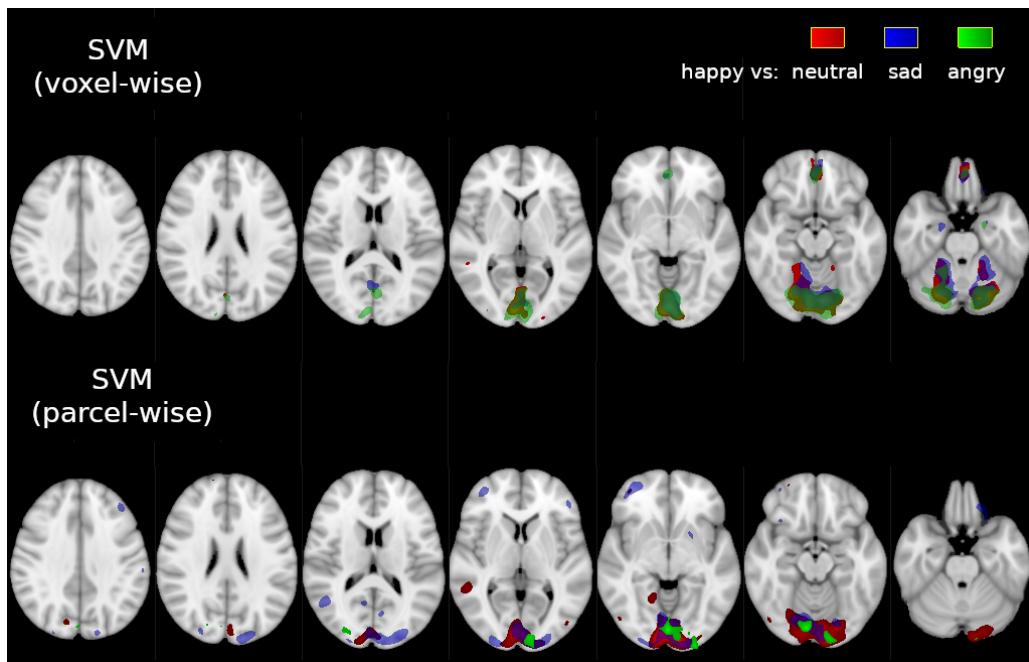


Figura 7.12: asdf.

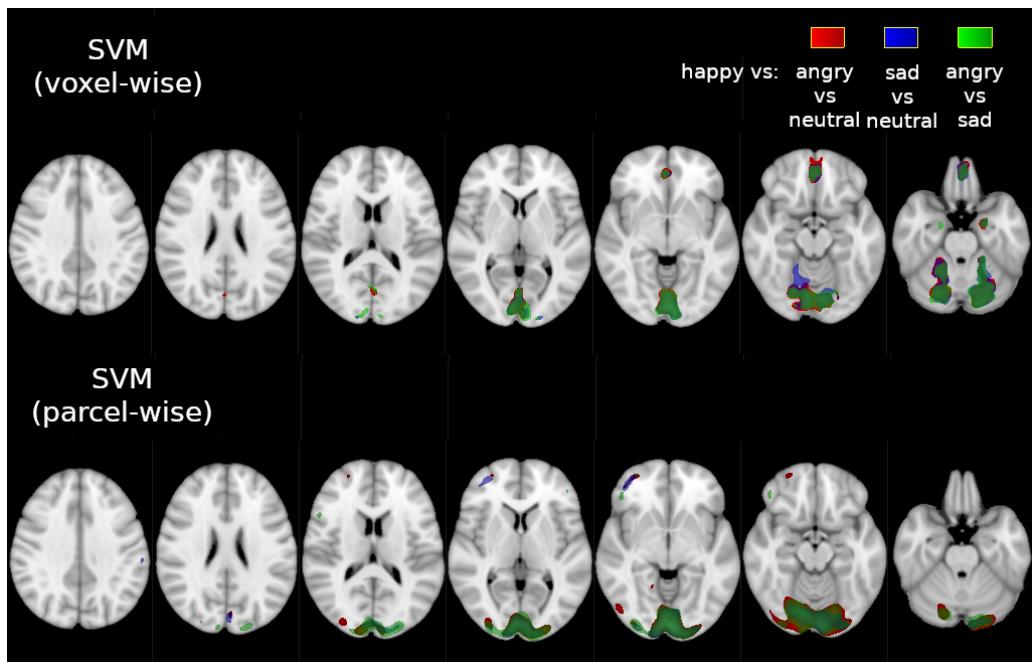


Figura 7.13: asdf.

# Capítulo 8

## Discusión

# **Capítulo 9**

## **Conclusiones**

# Apéndice 1: prueba de la función de verosimilitud de la regresión lineal simple

$$\begin{aligned}
\ln(\mathcal{L}(\mathbf{y}, \theta)) &= \ln \left( \prod_{i=1}^n \mathcal{N}(y_i | \mu, \sigma^2) \right) = \sum_{i=1}^n \ln \left[ \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{y_i - \mu}{\sigma} \right)^2} \right] \\
&= \sum_{i=1}^n \ln \left( \frac{1}{\sigma \sqrt{2\pi}} \right) + \sum_{i=1}^n \ln \left( e^{-\frac{1}{2} \left( \frac{y_i - \mu}{\sigma} \right)^2} \right) \\
&= -n \ln(\sigma \sqrt{2\pi}) + \sum_{i=1}^n -\frac{1}{2} \left( \frac{y_i - \mu}{\sigma} \right)^2 \\
&= -n \ln((2\pi\sigma^2)^{1/2}) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \\
&= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \underbrace{x_i \theta}_{\mu} \right)^2
\end{aligned}$$

## Apéndice 2: prueba de $\hat{\theta}$ por MLE para regresión lineal simple (ecuación normal)

$$\hat{\theta} = \arg_{\theta} \max \ln(\mathcal{L}) = \left\{ \theta : 0 = \frac{\partial \ln(\mathcal{L})}{\partial \theta} \right\}$$

$$0 = -\frac{1}{2\sigma^2} \sum_{i=1}^n 2(y_i - x_i\theta)(-x_i)$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i y_i - x_i^2 \theta)$$

$$= \sum_{i=1}^n x_i y_i - \theta \sum_{i=1}^n x_i^2$$

$$\hat{\theta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{x} \cdot \mathbf{x}} = \frac{Cov(x, y)}{Var(x)}$$

# Capítulo 10

## Apéndice 3: prueba de $\hat{\sigma}^2$ por MLE para regresión lineal simple

$$\begin{aligned}\hat{\sigma}^2 &= \arg_{\sigma^2} \max \ln(\mathcal{L}) = \left\{ \sigma^2 : 0 = \frac{\partial \ln(\mathcal{L})}{\partial \sigma^2} \right\} \\ 0 &= -\frac{n}{2} \left( \frac{1}{2\pi\sigma^2} \right) (2\pi) + \frac{1}{\sigma^4} \left( \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2 \right) \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 \\ &= \frac{1}{2\sigma^2} \left( \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 - n \right) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 - n \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2 = \text{Var}(y)\end{aligned}$$

# Apéndice 1: código fuente

Listing 10.1: Modelos predictivos de un solo voxel vs multivoxel. El siguiente programa en R fue utilizado para generar las figuras 2.10 y 2.9. Además se incluye una breve demostración del clasificador SVM.

```
1 library(ggplot2)
2 theme_set(theme_gray(base_size = 18))
3 library(cowplot)
4 library(e1071) # SVM classifier
5
6 set.seed(111)
7 N <- 40
8 CONDITIONS <- c("A", "B")
9
10 ## uniformly sample N points within (0,1)x(0,1)
11 data <- data.frame(replicate(2, runif(N, min = -.5, max = .5)))
12 colnames(data) <- c("voxel_1", "voxel_2")
13
14 ## divide space into 2 condition regions according to some boundary relation
15 label <- function(p) {
16   ## showcase nonlinear capabilities: circle of area .5 centered at (0, 0)
17   diameter <- .5 / pi
18   if ((p["voxel_1"])**2 > diameter - (p["voxel_2"])**2) {
19     CONDITIONS[1]
20   } else {
21     CONDITIONS[2]
22   }
23 }
24 data <- cbind(data, cond = apply(data, 1, label)) # label each point
25 ## reorder according to label, for visual convenience
26 data <- data[with(data, order(cond)), ]
27
28 ## plot single-voxel models
29 lm_plot <- function(data, xname, yname, remove_ytext = FALSE) {
30   model <- summary(lm(paste0(yname, "~", xname),
31                         data))
32   beta1 <- model$coefficients[paste0(xname, "B"), "Estimate"]
33   beta0 <- model$coefficients["(Intercept)", "Estimate"]
34   p <- model$coefficients[paste0(xname, "B"), "Pr(>|t|)"]
35   plot <- ggplot(data, aes(x = data[, xname],
36                         y = data[, yname],
37                         color = data[, xname])) +
38   labs(x = "Condición", y = "Intensidad") +
39   geom_violin() +
40   geom_jitter(width = .1, size = 3) +
41   geom_abline(slope = beta1,
42               intercept = beta0) +
43   geom_text(color = "black",
44             angle = atan(beta1) * (180 / pi) * 3.75,
45             aes(x = 1.5,
46                 y = 0,
47                 label = paste0("y=",
48                               round(beta1, 2),
49                               "x ",
50                               round(beta0, 2),
51                               ", p=",
52                               round(p, 2))),
53             size = 6) +
54   theme(legend.position = "none")
55   if (remove_ytext) {
56     plot <- plot + theme(axis.title.y = element_blank(),
57                           axis.text.y = element_blank(),
```

```

58         axis.ticks.y = element_blank())
59     }
60     shapiro.test(data[data$cond == CONDITIONS[1], as.character(yname)])
61     shapiro.test(data[data$cond == CONDITIONS[2], as.character(yname)])
62   plot
63 }
64
65 svg("./lm-vs-multivariate-1.svg")
66 cowplot::plot_grid(lm_plot(data, xname = "cond", yname = "voxel_1"), # no effect, p = .795
67   lm_plot(data, xname = "cond", yname = "voxel_2", TRUE), # no effect, p = .546
68   labels = c("Voxel 1", "Voxel 2"),
69   label_x = .5)
70 dev.off()
71
72 ## summary(lm(formula = cond ~ voxel_1:voxel_2, data = data)) # no interaction
73
74 svg("./lm-vs-multivariate-2.svg")
75 ggplot(data, aes(x = voxel_1, y = voxel_2, color = cond, shape = cond)) +
76   geom_point(size = 4) +
77   labs(x = "Intensidad (voxel 1)",
78        y = "Intensidad (voxel 2)",
79        color = "Condition") +
80   scale_color_discrete("") +
81   scale_shape_manual("", values = c(16, 17))
82 dev.off()
83
84 indices <- sample(1:nrow(data), N/2)
85 model3 <- svm(cond ~ voxel_2 + voxel_1,
86   data = data[indices, ],
87   kernel = "radial",
88   cost = 1000,
89   gamma = 0.01)
90 predict(model3, data[-indices, ])
91 svg("./lm-vs-multivariate-3.svg")
92 plot(model3, data)
93 dev.off()

```

# Referencias

- Allman, J.M. et al., 2001. The anterior cingulate cortex: The evolution of an interface between emotion and cognition. *Annals of the New York Academy of Sciences*, 935(1), pp.107–117.
- American Psychiatric Association, 2013. *Diagnostic and statistical manual of mental disorders (DSM-5)*, American Psychiatric Pub.
- Barrett, L.F., 2006. Are emotions natural kinds? *Perspectives on psychological science*, 1(1), pp.28–58.
- Barrett, L.F., Wager, T.D., 2006. The structure of emotion: Evidence from neuroimaging studies. *Current Directions in Psychological Science*, 15(2), pp.79–83.
- Baucom, L.B. et al., 2012. Decoding the neural representation of affective states. *Neuroimage*, 59(1), pp.718–727.
- Bechara, A., Damasio, A.R., Damasio, H., 2001. Insensitivity to future consequences following damage to human prefrontal cortex. *The Science of Mental Health: Personality and personality disorder*, 50, p.287.
- Bechara, A., Damasio, H., Damasio, A.R., 2000. Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex*, 10(3), pp.295–307.
- Bedetti, C. et al., 2019. *Cbedetti/dcm2bids 2.1.4*, Zenodo. Available at: <https://doi.org/10.5281/zenodo.2628725>.
- Bentham, J., 1781. *An introduction to the principles of morals and legislation*, McMaster University Archive for the History of Economic Thought.
- Berridge, K.C., Kringelbach, M.L., 2008. Affective neuroscience of pleasure: Reward in humans and animals. *Psychopharmacology*, 199(3), pp.457–480.
- Berrios, G.E., 1993. European views on personality disorders: A conceptual history. *Comprehensive Psychiatry*, 34(1), pp.14–30.
- Bush, G., Luu, P., Posner, M.I., 2000. Cognitive and emotional influences in anterior cingulate cortex. *Trends in cognitive sciences*, 4(6), pp.215–222.
- Calder, A.J., Lawrence, A.D., Young, A.W., 2001. Neuropsychology of fear and loathing. *Nature Reviews Neuroscience*, 2(5), p.352.
- Cannon, W.B., 1927. The james-lange theory of emotions: A critical examination and an alternative theory. *The American journal of psychology*, 39(1/4), pp.106–124.
- Celeghin, A. et al., 2017. Basic emotions in human neuroscience: Neuroimaging and beyond. *Frontiers in Psychology*, 8, p.1432.
- Cesario, J., Johnson, D.J., Eisthen, H.L., 2020. Your brain is not an onion with a tiny reptile inside. *Current Directions in Psychological Science*, p.0963721420917687.
- Chalmers, D.J., 1995. Facing up to the problem of consciousness. *Journal of consciousness studies*, 2(3),

pp.200–219.

Chang, L.J. et al., 2015. A sensitive and specific neural signature for picture-induced negative affect. *PLoS biology*, 13(6), p.e1002180.

Chialvo, D.R., 2010. Emergent complex neural dynamics. *Nature physics*, 6(10), pp.744–750.

Chikazoe, J. et al., 2014. Population coding of affect across stimuli, modalities and individuals. *Nature neuroscience*, 17(8), p.1114.

Conley, M.I. et al., 2018. The racially diverse affective expression (radiate) face stimulus set. *Psychiatry research*, 270, pp.1059–1067.

Cox, R. et al., 2004. A (sort of) new image data format standard: NIfTI-1: WE 150. *Neuroimage*, 22.

Damasio, A., Carvalho, G.B., 2013. The nature of feelings: Evolutionary and neurobiological origins. *Nature reviews neuroscience*, 14(2), p.143.

Damasio, A.R., 1999. *The feeling of what happens: Body and emotion in the making of consciousness*, Houghton Mifflin Harcourt.

Damasio, A.R., 1996. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 351(1346), pp.1413–1420.

Darwin, C., 1872. *The expression of the emotions in man and animals*, John Murray.

Dawkins, R., 1976. *The selfish gene*, Oxford university press.

Donath, J., 2007. Signals, cues and meaning. *Unpublished Manuscript. Massachusetts Institute of Technology, Cambridge, MA*. Available at: <http://smg.media.mit.edu/classes/IdentitySignals06/SignalingDraft.pdf>.

Ekman, P., 2003. Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, 1000(1), pp.205–221.

Ekman, P., 1976. Pictures of facial affect. *Consulting Psychologists Press*.

Ekman, P.E., Davidson, R.J., 1994. *The nature of emotion: Fundamental questions*, Oxford University Press.

Ellis, H.D. et al., 1997. Reduced autonomic responses to faces in capgras delusion. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1384), pp.1085–1092.

Ethofer, T. et al., 2009. Decoding of emotional information in voice-sensitive cortices. *Current Biology*, 19(12), pp.1028–1033.

Eysenck, M.W., 2004. *Psychology: An international perspective*, Taylor & Francis.

Fessler, D.M.T., Gervais, M., 2010. From whence the captains of our lives: Ultimate and phylogenetic perspectives on emotions in humans and other primates. In P. M. Kappeler, J. Silk, eds. *Mind the gap: Tracing the origins of human universals*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 261–280.

Fisher, R.A., 1922. The goodness of fit of regression formulae, and the distribution of regression coefficients. *Journal of the Royal Statistical Society*, 85(4), pp.597–612.

- Fonov, V. et al., 2011. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*, 54(1), pp.313–327.
- Fonov, V.S. et al., 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, (47), p.S102.
- Fridlund, A.J., 2014. *Human facial expression: An evolutionary view*, Academic Press.
- Friesen, C.K., Kingstone, A., 1998. The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic bulletin & review*, 5(3), pp.490–495.
- Friston, K., 2010. The free-energy principle: A unified brain theory? *Nature reviews neuroscience*, 11(2), p.127.
- Friston, K.J. et al., 1991. Comparing functional (pet) images: The assessment of significant change. *Journal of Cerebral Blood Flow & Metabolism*, 11(4), pp.690–699.
- Friston, K.J. et al., 1994. Statistical parametric maps in functional imaging: A general linear approach. *Human brain mapping*, 2(4), pp.189–210.
- Gauss, C.F., 1823. *Theoria combinationis observationum erroribus minimis obnoxiae*, H. Dieterich.
- Gazzaniga, M., Ivry, R.B., 2013. *Cognitive neuroscience: The biology of the mind: Fourth international student edition*, WW Norton.
- Gerlach, W., Stern, O., 1922a. Das magnetische moment des silberatoms. *Zeitschrift für Physik*, 9(1), pp.353–355.
- Gerlach, W., Stern, O., 1922b. Der experimentelle nachweis der richtungsquantelung im magnetfeld. *Zeitschrift für Physik*, 9(1), pp.349–352.
- Gorgolewski, K.J. et al., 2016. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, 3(1), pp.1–9.
- Grossman, S. et al., 2019. Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. *Nature communications*, 10(1), pp.1–13.
- Gu, X. et al., 2013. Anterior insular cortex and emotional awareness. *Journal of Comparative Neurology*, 521(15), pp.3371–3388.
- Guillory, S.A., Bujarski, K.A., 2014. Exploring emotions using invasive methods: Review of 60 years of human intracranial electrophysiology. *Social cognitive and affective neuroscience*, 9(12), pp.1880–1889.
- Gulban, O.F. et al., 2019. *Poldracklab/pydeface: V2.0.0*, Zenodo. Available at: <https://doi.org/10.5281/zenodo.3524401>.
- Haist, F., Anzures, G., 2017. Functional development of the brain's face-processing system. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(1-2), p.e1423.
- Hamann, S., 2012. Mapping discrete and dimensional emotions onto the brain: Controversies and consensus. *Trends in cognitive sciences*, 16(9), pp.458–466.
- Harrison, N.A. et al., 2010. The embodiment of emotional feelings in the brain. *Journal of Neuroscience*, 30(38), pp.12878–12884.

- Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6), pp.223–233.
- Haynes, J.-D., Rees, G., 2005. Predicting the stream of consciousness from activity in human visual cortex. *Current Biology*, 15(14), pp.1301–1307.
- Hernández-Pérez, R., Concha, L., Cuaya, L.V., 2018. Decoding human emotional faces in the dog's brain. *bioRxiv*, p.134080.
- Huettel, S.A. et al., 2009. *Functional magnetic resonance imaging* 2nd ed., Sinauer Associates Sunderland, MA.
- Hurlemann, R. et al., 2007. Amygdala control of emotion-induced forgetting and remembering: Evidence from urbach-wiethe disease. *Neuropsychologia*, 45(5), pp.877–884.
- Isen, A.M., Patrick, R., 1983. The effect of positive feelings on risk taking: When the chips are down. *Organizational behavior and human performance*, 31(2), pp.194–202.
- James, W., 1884. What is an emotion? *Mind*, 9(34), pp.188–205.
- Jenkinson, M. et al., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2), pp.825–841.
- Jenkinson, M. et al., 2012. Fsl. *Neuroimage*, 62(2), pp.782–790.
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5(2), pp.143–156.
- Jevons, W.S., 1879. *The theory of political economy*, Macmillan.
- Jimura, K., Poldrack, R.A., 2012. Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia*, 50(4), pp.544–552.
- Kassam, K.S. et al., 2013. Identifying emotions on the basis of neural activation. *PloS one*, 8(6), p.e66032.
- Klüver, H., Bucy, P.C., 1937. "Psychic blindness." and other symptoms following bilateral temporal lobectomy in rhesus monkeys. *American Journal of Physiology*.
- Kober, H. et al., 2008. Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *Neuroimage*, 42(2), pp.998–1031.
- Körding, K.P., Wolpert, D.M., 2006. Bayesian decision theory in sensorimotor control. *Trends in cognitive sciences*, 10(7), pp.319–326.
- Kotz, S.A. et al., 2012. Predicting vocal emotion expressions from the human brain. *Human Brain Mapping*, 34(8), pp.1971–1981.
- Kragel, P.A., LaBar, K.S., 2014. Advancing emotion theory with multivariate pattern classification. *Emotion Review*, 6(2), pp.160–174.
- Kragel, P.A., LaBar, K.S., 2016. Decoding the nature of emotion in the brain. *Trends in cognitive sciences*, 20(6), pp.444–455.
- Kragel, P.A., LaBar, K.S., 2015. Multivariate neural biomarkers of emotional states are categorically

distinct. *Social cognitive and affective neuroscience*, 10(11), pp.1437–1448.

Laidre, M.E., Johnstone, R.A., 2013. Animal signals. *Current Biology*, 23(18), pp.R829–R833.

LeDoux, J.E., Damasio, A.R., 2013. In E. R. Kandel et al., eds. *Principles of neural science*. McGraw-hill New York.

Legendre, A.M., 1805. *Nouvelles méthodes pour la détermination des orbites des comètes*, F. Didot.

Lewis, T., 2013. Oldest fossils reveal when apes & monkeys first diverged. Available at: <https://www.livescience.com/32029-oldest-monkey-fossil-found.html>.

Lewis-Peacock, J.A., Norman, K.A., 2013. Multi-voxel pattern analysis of fMRI data. *The cognitive neurosciences*, pp.911–920.

Li, X. et al., 2016. The first step for neuroimaging data analysis: DICOM to nifti conversion. *Journal of neuroscience methods*, 264, pp.47–56.

Lindquist, K.A., Barrett, L.F., 2012. A functional architecture of the human brain: Emerging insights from the science of emotion. *Trends in cognitive sciences*, 16(11), pp.533–540.

Lindquist, K.A. et al., 2012. The brain basis of emotion: A meta-analytic review. *Behavioral and brain sciences*, 35(3), pp.121–143.

Liu, J. et al., 2014. Seeing jesus in toast: Neural and behavioral correlates of face pareidolia. *Cortex*, 53, pp.60–77.

Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. *Nature*, 453(7197), p.869.

Logothetis, N.K. et al., 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), pp.150–157.

MacLean, P.D., 1949. Psychosomatic disease and the "visceral brain"; recent developments bearing on the papez theory of emotion. *Psychosomatic medicine*.

Mahmoudi, A. et al., 2012. Multivoxel pattern analysis for fmri data: A review. *Computational and mathematical methods in medicine*, 2012.

Manjón, J.V., Coupé, P., 2016. VolBrain: An online mri brain volumetry system. *Frontiers in neuroinformatics*, 10, p.30.

Marlowe, W.B., Mancall, E.L., Thomas, J.J., 1975. Complete Klüver-Bucy syndrome in man. *Cortex*, 11(1), pp.53–59.

Michel, M. et al., 2019. Opportunities and challenges for a maturing science of consciousness. *Nature human behaviour*, 3(2), p.104.

Millon, T., Davis, R.O., 1996. *Disorders of personality: DSM-iv and beyond*, John Wiley & Sons.

Mitchell, R.L., Phillips, L.H., 2015. The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia*, 70, pp.1–10.

Murphy, F.C., Nimmo-Smith, I., Lawrence, A.D., 2003. Functional neuroanatomy of emotions: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience*, 3(3), pp.207–233.

- Mustra, M., Delac, K., Grgic, M., 2008. Overview of the dicom standard. In *2008 50th international symposium elmar*. IEEE, pp. 39–44.
- Nelder, J.A., Wedderburn, R.W., 1972. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3), pp.370–384.
- Newen, A., De Bruin, L., Gallagher, S., 2018. *The oxford handbook of 4E cognition*, Oxford University Press.
- Norman, K.A. et al., 2006. Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in cognitive sciences*, 10(9), pp.424–430.
- Ogawa, S., Lee, T.-M., 1990. Magnetic resonance imaging of blood vessels at high fields: In vivo and in vitro measurements and image simulation. *Magnetic resonance in medicine*, 16(1), pp.9–18.
- Ogawa, S. et al., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *proceedings of the National Academy of Sciences*, 87(24), pp.9868–9872.
- Ogawa, S. et al., 1990. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic resonance in medicine*, 14(1), pp.68–78.
- Ogawa, S. et al., 1992. Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, 89(13), pp.5951–5955.
- Oizumi, M., Albantakis, L., Tononi, G., 2014. From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS computational biology*, 10(5), p.e1003588.
- Papez, J.W., 1937. A proposed mechanism of emotion. *Archives of Neurology & Psychiatry*, 38(4), pp.725–743.
- Pauli, W., 1924. Über den einfluß der geschwindigkeitsabhängigkeit der elektronenmasse auf den zeeman-effekt. *Zeitschrift für Physik*, 31(1), pp.373–385.
- Pauli, W., 1927. Zur quantenmechanik des magnetischen elektrons. In *Wolfgang pauli*. Springer, pp. 282–305.
- Pauling, L., Coryell, C.D., 1936. The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences*, 22(4), pp.210–216.
- Pearson, K., 1903. The law of ancestral heredity. *Biometrika*, 2(2), pp.211–228.
- Peelen, M.V., Atkinson, A.P., Vuilleumier, P., 2010. Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience*, 30(30), pp.10127–10134.
- Pessoa, L., Padmala, S., 2007. Decoding near-threshold perception of fear from distributed single-trial brain activation. *Cerebral cortex*, 17(3), pp.691–701.
- Phan, K.L. et al., 2002. Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, 16(2), pp.331–348.
- Polyn, S.M. et al., 2005. Category-specific cortical activity precedes retrieval during memory search. *Science*, 310(5756), pp.1963–1966.

- Porter, S., Ten Brinke, L., 2008. Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological science*, 19(5), pp.508–514.
- Preston, T., 1898. Radiation phenomena in a strong magnetic field:... In R. Dublin Soc.
- Quiroga, R.Q. et al., 2005. Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045), pp.1102–1107.
- Ramachandran, V., 1998. Consciousness and body image: Lessons from phantom limbs, capgras syndrome and pain asymbolia. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1377), pp.1851–1859.
- Reggev, N. et al., 2020. Human face-selective cortex does not distinguish between members of a racial outgroup. *Eneuro*, 7(3).
- Rissman, J., Wagner, A.D., 2012. Distributed representations in memory: Insights from functional brain imaging. *Annual review of psychology*, 63, pp.101–128.
- Rohrer, J.D. et al., 2011. Clinical and neuroanatomical signatures of tissue pathology in frontotemporal lobar degeneration. *Brain*, 134(9), pp.2565–2581.
- Rolls, E.T., Grabenhorst, F., Franco, L., 2009. Prediction of subjective affective state from brain activations. *Journal of Neurophysiology*, 101(3), pp.1294–1308.
- Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage*, 52(3), pp.1059–1069.
- Russell, J.A., Dols, J.M.F., 1997. *The psychology of facial expression*, Cambridge university press Cambridge.
- Saarimäki, H. et al., 2015. Discrete neural signatures of basic emotions. *Cerebral cortex*, 26(6), pp.2563–2573.
- Said, C.P. et al., 2010. Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of vision*, 10(5), pp.11–11.
- Sander, D., Grafman, J., Zalla, T., 2003. The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, 14(4), pp.303–316.
- Schacter, D.L., Gilbert, D.T., Wegner, D.M., 2011. *Psychology (2nd edition)*, New York: Worth; Worth.
- Schirmer, A., Adolphs, R., 2017. Emotion perception from face, voice, and touch: Comparisons and convergence. *Trends in cognitive sciences*, 21(3), pp.216–228.
- Schmahmann, J.D., Sherman, J.C., 1997. Cerebellar cognitive affective syndrome. *International review of neurobiology*, 41, pp.433–440.
- Schwarz, N., 1990. Feelings as information: Informational and motivational functions of affective states. In T. E. Higgins, R. M. Sorrentino, eds. *Handbook of motivation and cognition: Foundations of social behavior*. Guilford Press, pp. 527–561.
- Schwarz, N., Clore, G.L., 2003. Mood as information: 20 years later. *Psychological inquiry*, 14(3-4), pp.296–303.

Scidraw, 2020. Neuron silhouette. Zenodo. Creative commons attribution 4.0 international. Available at: <https://doi.org/10.5281/zenodo.3925927>.

Seth, A.K., 2013. Interoceptive inference, emotion, and the embodied self. *Trends in cognitive sciences*, 17(11), pp.565–573.

Seth, A.K., 2010. The grand challenge of consciousness. *Frontiers in psychology*, 1, p.5.

Shamay-Tsoory, S.G., Aharon-Peretz, J., Perry, D., 2009. Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, 132(3), pp.617–627.

Sherrington, C.S., 1900. Experiments on the value of vascular and visceral factors for the genesis of emotion. *Proceedings of the Royal Society of London*, 66(424-433), pp.390–403.

Shinkareva, S.V. et al., 2014. Representations of modality-specific affective processing for visual and auditory stimuli derived from functional magnetic resonance imaging data. *Human brain mapping*, 35(7), pp.3558–3568.

Simon, D. et al., 2008. Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *PAIN*, 135(1-2), pp.55–64.

Simon, D. et al., 2006. Brain responses to dynamic facial expressions of pain. *Pain*, 126(1-3), pp.309–318.

Sitaram, R. et al., 2011. Real-time support vector classification and feedback of multiple emotional brain states. *Neuroimage*, 56(2), pp.753–765.

Skerry, A.E., Saxe, R., 2015. Neural representations of emotion are organized around abstract event features. *Current biology*, 25(15), pp.1945–1954.

Slaby, J., 2014. Emotions and the extended mind. *Collective emotions*, pp.32–46.

Soon, C.S. et al., 2008. Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, 11(5), p.543.

Stringer, C. et al., 2019. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437), p.eaav7893.

Sutton, R.S., Barto, A.G., 2018. *Reinforcement learning: An introduction*,

Timbergen, N., 1963. On aims and methods of ethology. *Zeitschrift für tierpsychologie*, 20(4), pp.410–433.

Tong, F., Pratte, M.S., 2012. Decoding patterns of human brain activity. *Annual review of psychology*, 63, pp.483–509.

Tononi, G., Koch, C., 2015. Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668), p.20140167.

Torralva, T. et al., 2009. A neuropsychological battery to detect specific executive and social cognitive impairments in early frontotemporal dementia. *Brain*, 132(5), pp.1299–1309.

Tottenham, N. et al., 2009. The nimstim set of facial expressions: Judgments from untrained research participants. *Psychiatry research*, 168(3), pp.242–249.

- Tracy, J.L., Randles, D., Steckler, C.M., 2015. The nonverbal communication of emotions. *Current opinion in behavioral sciences*, 3, pp.25–30.
- Triarhou, L.C., 2008. Centenary of christfried jakob's discovery of the visceral brain: An unheeded precedence in affective neuroscience. *Neuroscience & Biobehavioral Reviews*, 32(5), pp.984–1000.
- Tsao, D.Y. et al., 2006. A cortical region consisting entirely of face-selective cells. *Science*, 311(5761), pp.670–674.
- Tsao, D.Y., Livingstone, M.S., 2008. Mechanisms of face perception. *Annu. Rev. Neurosci.*, 31, pp.411–437.
- Tsuchiya, N., Adolphs, R., 2007. Emotion and consciousness. *Trends in cognitive sciences*, 11(4), pp.158–167.
- Uhlenbeck, G.E., Goudsmit, S., 1925. Ersetzung der hypothese vom unmechanischen zwang durch eineforderung bezüglich des inneren verhaltens jedes einzelnen elektrons. *Naturwissenschaften*, 13(47), pp.953–954.
- Uludag, K., Dubowitz, D.J., Buxton, R.B., 2005. Basic principles of functional MRI. *Clinical MRI. Elsevier, San Diego*, pp.249–287.
- Vytal, K., Hamann, S., 2010. Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis. *Journal of cognitive neuroscience*, 22(12), pp.2864–2885.
- Wager, T.D. et al., 2008. The neuroimaging of emotion. *Handbook of emotions*, 3, pp.249–271.
- Weigrzyn, M. et al., 2015. Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex*, 69, pp.131–140.
- Weiskrantz, L., 1956. Behavioral changes associated with ablation of the amygdaloid complex in monkeys. *Journal of comparative and physiological psychology*, 49(4), p.381.
- Wilson, E.O., 2012. *The social conquest of earth*, WW Norton & Company.
- Woolrich, M.W. et al., 2001. Temporal autocorrelation in univariate linear modeling of fmri data. *Neuroimage*, 14(6), pp.1370–1386.
- World Health Organization, 2004. *International statistical classification of diseases and related health problems* 10th ed., World Health Organization.
- Worsley, K.J. et al., 1992. A three-dimensional statistical analysis for cbf activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, 12(6), pp.900–918.
- Yule, G.U., 1897. On the theory of correlation. *Journal of the Royal Statistical Society*, 60(4), pp.812–854.
- Zaki, J., Ochsner, K.N., 2012. The neuroscience of empathy: Progress, pitfalls and promise. *Nature neuroscience*, 15(5), p.675.