



THE EMOTIONAL SELF: A RELATIONAL VIEW

El sujeto emocional: un punto de vista relacional

TESIS DOCTORAL

presentada por

MARTA CABRERA MIQUEL

dirigida por

DR. JOSEP E. CORBÍ FERNÁNDEZ DE IBARRA

DR. JORDI VALOR ABAD

Programa de Doctorado en Lógica y Filosofía de la Ciencia

Septiembre 2022

Facultat de Filosofia i Ciències de l'Educació

Universitat de València

Agradecimientos

Tengo la suerte de haber sido acompañada, guiada y cuidada por muchas personas a lo largo de estos cinco años de doctorado. A Josep Corbí le agradezco enormemente el tiempo que ha dedicado a leer mis textos y a conversar conmigo, así como su manera atenta y sensible de orientarme en mi trabajo. La claridad y profundidad de su pensamiento y su convicción de que la filosofía ha de estar vinculada a la vida han sido una fuente de inspiración muy importante para mí y para esta tesis. A Jordi Valor le doy las gracias por el apoyo constante que lleva ofreciéndome desde antes del doctorado y por la rigurosidad y el cuidado con el que ha comentado mi trabajo. Gracias a Jordi, sentí que podía haber un hueco para mí en el ámbito de la filosofía académica y decidí hacer esta tesis.

Estoy muy agradecida también a todos los miembros del VLC Philosophy Lab por haberme proporcionado un espacio en el que desarrollarme como investigadora. Formar parte de este grupo diverso y acogedor ha sido fundamental para mí a lo largo de estos años. En especial, le agradezco a Carlos Moya que me supervisara durante el primer año del doctorado y me pusiera en contacto con el fascinante mundo de la filosofía de la acción y a Pablo Rychter y a Chon Tejedor, que compartieran conmigo su docencia y me ayudaran durante el proceso. Al grupo de doctorandas y doctorandos le debo mucho también: Virginia Ballesteros, Ana Batalla, Javier Castellote, Joan Gimeno y Carmen Martínez me han inspirado con sus respectivos trabajos y han sido un apoyo muy importante durante la, a ratos, compleja etapa predoctoral. La generosidad de Saúl Pérez-González merece un agradecimiento aparte: sin su paciencia, sus explicaciones y sus ánimos mi doctorado habría sido muy distinto.

A finales de 2018 tuve la oportunidad de realizar una estancia en la Universidad de Manchester bajo la supervisión de Tom Smith y, a principios de 2022, otra en la Universidad de Bolonia junto a Pia Campeggiani. Ambos me ayudaron mucho a mejorar la tesis a través de sus sugerencias y comentarios y me permitieron entrar en contacto con un gran número de investigadoras e investigadores especializados en filosofía de las emociones. En la Universidad de Manchester, además, tuve la suerte de conocer a Jansan Favazzo, Pilar López-Cantero, Benedetta Magro y Carlo Rainieri, quienes me ayudaron a entender cómo quería vivir la experiencia del doctorado.

Quiero dar las gracias, además, a las instituciones que me han apoyado económicamente. Sin la ayuda ACIF de la Generalitat Valenciana y la ayuda BES-2017-

081537 financiada por MCIN/AEI/10.13039/501100011033 y “FSE Invierte en tu futuro” no habría podido realizar esta tesis y, mucho menos, independizarme mientras la hacía. A los responsables de los proyectos “Self-Knowledge, Moral Responsibility and Authenticity” (FFI2016-75323-P) y “Eliminativism, Fictionalism, and Expressivism: The Possibility of a Negative Metaphysical Verdict about a Discourse D” (PID2019-106420GA-I00) les agradezco también su generosidad a la hora de financiar las actividades formativas en las que he participado así como la revisión de los artículos de la tesis.

Para acabar, quiero expresar mi más profundo agradecimiento a todas las personas que forman parte de mi red cotidiana y que han sido la condición de posibilidad de esta tesis. En especial, le agradezco a mi madre su incansable escucha; a Violeta, su presencia incondicional y a Aída, Sandra y Paula, la reflexión colectiva y el cuidado mutuo que dotan de sentido al trabajo que he realizado.

Índice

Resumen en castellano	6
Summary in English	38
Anexo 1: <i>On the Intentionality of Shame and Pride</i>	68
1. The Distinction	70
2. The Correctness Argument (CA)	73
3. The Phenomenological Argument (PA)	77
4. The Grammatical Argument (GA)	81
5. Conclusions	84
Anexo 2: <i>Duchenne Smiles are Actions not Mere Happenings: Lessons from the Debate on Expressive Action</i>	87
1. Facial and bodily expressions of emotion	89
2. Simple expressive actions	92
3. Facial and bodily expressions of emotion versus simple expressive actions	97
4. Conclusions	103
Anexo 3: <i>¿Qué es una emoción?</i>	105
1. Introducción	107
2. Aclaraciones iniciales	108
2.1. El enfoque	108
2.2. El fenómeno	109
3. Emociones: fenómenos intencionales, evaluativos y sentidos	113
3.1. El aspecto intencional de las emociones	114
3.2. El aspecto evaluativo de las emociones	115
3.3. El aspecto sentido de las emociones	116
4. Teorías de las emociones	118
4.1. Teorías cognitivistas	118
4.2. La teoría perceptiva	125
4.3. Teorías del sentir	129
5. El problema de la relación entre emociones y propiedades evaluativas	139
5.1. Las propiedades evaluativas como independientes de las emociones	140
5.2. La relación de dependencia mutua entre emociones y propiedades evaluativas	142
6. Conclusión	149

Resumen en castellano

1. Introducción

Siempre he sentido un interés especial por el aspecto íntimo de las emociones. Aunque las emociones son experiencias que nos conectan de manera profunda con los demás, hay algo en ellas que las convierte, al mismo tiempo, en las experiencias más íntimas. El sentido en el que entiendo esto último se aleja de las nociones de privacidad o interioridad propias del debate sobre la asimetría entre la primera y la tercera persona. La noción de intimidad a la que aludo está vinculada, más bien, a la experiencia de sentirse vulnerable. Como es bien sabido, sólo podemos sentir una emoción por algo que nos importa y, en este sentido, toda experiencia emocional es una experiencia en la que estamos expuestos, esto es, en la que nos jugamos algo. Desde esta perspectiva, profundizar en el aspecto íntimo de las emociones supone ahondar en la manera en la que estamos vinculados a las cosas que nos importan.

Explorar la manera en la que nuestro yo es vulnerable a lo que le rodea resulta una tarea fundamental para cualquiera que piense que conocerse a una misma es un requisito básico para una vida con sentido. Al menos, este fue mi punto de partida cuando decidí hacer una tesis doctoral en filosofía de las emociones. Aunque las emociones están estrechamente relacionadas con un gran número de elementos – el cuerpo, el mundo, los demás, los valores, etc. – la cuestión del yo que se expresa y se expone en ellas me parecía tener una cierta prioridad sobre las demás cuestiones. Desde una perspectiva algo ingenua, me preguntaba cómo era posible estudiar la manera en la que nuestras emociones se relacionan con los demás o con el mundo sin examinar primero la manera en la que el yo está involucrado en ellas. Me parecía que el aspecto más básico de nuestras emociones es que, en ellas, el protagonismo del yo es innegable: respondo emocionalmente a un objeto porque tal objeto tiene algo que ver *conmigo*, esto es, *me concierne*. Y esto es así independientemente de cuál sea el objeto de la emoción, de la proporcionalidad de la respuesta, de la verdad o la falsedad de las creencias involucradas, etc. Esta convicción me condujo a sumergirme en la cuestión del yo en las emociones entendida como una cuestión previa y, en algún sentido, separada del conjunto de elementos mencionados más arriba.

A lo largo del trabajo que he realizado, he llegado, no obstante, a cuestionarme que podamos aproximarnos al yo como una realidad separada del cuerpo, de los otros y del mundo. Paradójicamente, mi interés por entender el aspecto íntimo de las emociones – donde la vulnerabilidad del yo habría de manifestarse de la manera más clara – me ha

llevado a la convicción de que lo específico del yo que se expresa en las emociones es su estrecha relación o profunda continuidad con el mundo, el cuerpo y los demás. En cierto sentido, en mis intentos por centrarme en el yo, he tenido que desandar el camino que pensaba que me llevaría hasta él para volver a una serie de elementos – que normalmente se entienden por oposición al yo – sin los cuales el yo no puede ser más que una mera simplificación teórica.

En este resumen, voy a comenzar explicando mi primer acercamiento a la filosofía de las emociones y a la cuestión del yo que se expresa en ellas. Como comentaré brevemente, el resultado de este primer acercamiento fue la convicción de que era necesario revisar la manera en la que suele concebirse al sujeto en el debate filosófico actual sobre las emociones. A continuación, expondré las tres tesis que defiendo en este trabajo e intentaré mostrar cómo, en su conjunto, pretenden poner en cuestión lo que llamaré “la concepción escindida del sujeto”. Después de esto, resumiré el contenido y las conclusiones más importantes de los tres artículos que componen esta tesis.

2. Primera aproximación al yo en las emociones

En mi primera aproximación al debate filosófico sobre las emociones, me fascinó el trabajo de Peter Goldie y, en concreto, su idea de que las emociones, en tanto que estados mentales, tienen un contenido deíctico que nos ubica, nos compromete emocionalmente con el mundo y nos prepara para la acción de un modo especial. Al hilo del ejemplo de John Perry (1979) en el que un profesor cree que tiene una reunión por la tarde pero sólo se levantará para asistir a la misma cuando crea, además, que *ahora* es por la tarde – aunque su creencia siga teniendo exactamente el mismo contenido antes y después de levantarse: “la reunión empieza por la tarde” –, Goldie introduce el siguiente ejemplo:

Analogously, imagine you are in a zoo, looking at a gorilla grimly loping from left to right in its cage. You are thinking of the gorilla as dangerous, but you do not feel fear, as it seems to be safely behind bars. Then you see that the door to the cage has been left wide open. Just for a moment, though, you fail to put the two thoughts—the gorilla is dangerous, the cage is open—together. Then, suddenly, you do put them together: now your way of thinking of the gorilla as dangerous is new; now it is dangerous in an emotionally relevant way *for you*. The earlier thought, naturally expressed as ‘That gorilla is dangerous’, differs in content from the new thought, although this new thought, thought with emotional feeling, might also be naturally expressed in the same words. Now, in feeling fear towards the gorilla you are emotionally engaged with the world,

and, typically, you are poised for action in a new way—poised for action out of the emotion (2000a, 61, énfasis añadido).

Las primeras veces que leí este ejemplo – sobre el que Goldie no dice nada más – tuve la intuición de que en él residía la clave para entender el aspecto íntimo de las emociones. Me daba la impresión de que Goldie estaba intentando captar la manera en la que las emociones, a diferencia de otros estados mentales, añaden algo así como “esto tiene que ver *conmigo*” a los pensamientos que reflejan nuestra interacción con el mundo y los demás. Este ejemplo parecía recoger la intuición de que las emociones, además de apuntar hacia el mundo, siempre hacen referencia a una misma y a sus preocupaciones: al sentir miedo, mi percepción de que *yo* estoy en peligro permea completamente la percepción de que el objeto que tengo delante es peligroso y parece explicar el hecho de que, sólo cuando siento la emoción, me pongo en movimiento – echando a correr, por ejemplo. Del mismo modo, sólo es posible sentir enfado o indignación si el comentario ofensivo que acabo de escuchar me concierne de algún modo, sentir vergüenza por mis acciones o las de otra persona sólo es inteligible si tales acciones tienen algo que ver conmigo, etc. Que mi enfado, mi indignación o mi vergüenza me lleven a actuar de cierta forma se explicaría, precisamente, por el hecho de que estoy involucrada de manera relevante en lo que está ocurriendo.¹

A partir de esta lectura del fragmento de Goldie, me interesé por cómo las emociones parecen ubicarnos, es decir, por cómo parecen hacer referencia a la posición en la que estamos así como a la manera en la que estamos involucrados en una situación. Desde este punto de vista, las emociones tienen un gran valor epistémico no sólo en relación a las cosas que nos rodean – el miedo me informa de que estoy ante algo peligroso; la humillación, de que estoy ante algo humillante, etc. – sino en relación a nosotros mismos: dónde estoy, en qué sentido me afecta lo que tengo delante, cuáles son mis posibilidades de respuesta, cuáles son mis tendencias y mi carácter, etc.

A medida que profundizaba en la literatura sobre la relación entre las emociones y el yo, fui advirtiendo que tanto los distintos debates que analizaba como mi propio punto de partida estaban atravesados teóricamente por una cierta concepción del yo o del sujeto. En concreto, por una *concepción escindida del sujeto* que se caracteriza por concebirlo a

¹ Por descontado, puedo estar equivocada con respecto a si un comentario es ofensivo, sobre si su contenido está relacionado conmigo, etc. Lo importante es que, si no pienso que el comentario, las acciones o cualquiera que sea el objeto de la emoción tienen algo que ver conmigo – de manera más o menos directa –, no es inteligible que yo sienta una emoción al respecto.

través de una serie de oposiciones: sujeto-mundo, sujeto-cuerpo y sujeto-otros. Lo que en un principio me había parecido un punto de partida neutral, me parecía ahora que se correspondía con un conjunto de presupuestos filosóficos sobre el yo que, examinados con atención, resultan problemáticos y condicionan cualquier tipo de reflexión sobre el aspecto íntimo de las emociones. Esta preocupación transformó mi proyecto inicial en un intento de ahondar en esta concepción escindida del sujeto a fin de mostrar sus problemas y sugerir la importancia de una concepción alternativa que nos permita aproximarnos al yo que se expresa en las emociones de una manera más plausible.

3. Crítica a la concepción escindida del sujeto

Para profundizar en la manera en la que el debate filosófico actual sobre las emociones concibe al sujeto y poner de manifiesto los problemas derivados de dicha concepción, he analizado tres importantes debates contemporáneos: el debate sobre las emociones reflexivas (Cabrera en prensa), el debate sobre las acciones expresivas (Cabrera 2022b) y el debate sobre la relación entre emociones y propiedades evaluativas (Cabrera 2022a). En lugar de ofrecer una imagen completa de la manera estándar en la que se entiende al yo en las emociones, mi estrategia ha consistido en defender en cada uno de estos artículos una tesis que pone de manifiesto los problemas generados por dicha concepción.

Tesis 1: La vergüenza y el orgullo no están necesariamente dirigidos al yo como objeto intencional de la emoción.

Tesis 2: Somos agentes, y no pacientes, de las expresiones faciales y corporales de las emociones.

Tesis 3: Los problemas a los que se enfrentan varias de las teorías actuales de las emociones parecen tener su origen en una determinada concepción de las propiedades evaluativas como independientes de las respuestas emocionales del sujeto, y no en la supuesta complejidad de las emociones.²

En la defensa que realicé de estas tres tesis se pone de manifiesto la tendencia que existe en el debate filosófico actual sobre las emociones a identificar al sujeto con su vida

² Como expongo más adelante, el objetivo principal de Cabrera (2022a) es presentar un estado de la cuestión sobre la investigación filosófica en emociones. Sólo en la última sección de este artículo hago una breve revisión de la concepción de las propiedades evaluativas dominante en el debate actual sobre las emociones y argumento a favor de la *Tesis 3*.

mental, entendida como el conjunto de procesos cognitivos que tienen lugar dentro de su cabeza, y a tratarlo como una realidad aislada que puede ser concebida al margen del mundo, de su cuerpo y de los demás. En otras palabras, la defensa de estas tres tesis pretende mostrar la manera en la que las escisiones

- sujeto-mundo
- sujeto-cuerpo
- sujeto-otros

atraviesan de manera sistemática la reflexión contemporánea sobre las emociones. Esta manera de concebir al sujeto constituye *una concepción escindida* porque presenta al sujeto como disociado de una serie de realidades que son fundamentales para la comprensión de su vida emocional. Como explico brevemente a continuación, cada uno de los debates que examino en esta tesis asume alguna de las diferentes versiones en las que se concreta la concepción escindida del sujeto.

En primer lugar, el debate sobre las emociones reflexivas presupone una visión artificial según la cual *sujeto y objeto pueden disociarse nítidamente (escisión sujeto-mundo)*. Según los argumentos que reviso en Cabrera (en prensa), una emoción ha de dirigirse intencionalmente, o bien al sujeto, o bien a un objeto distinto del sujeto, dependiendo de cuál de estos dos elementos ejemplifique la propiedad evaluativa asociada a la emoción. Esta idea ha llevado a varios autores a concluir que el orgullo o la vergüenza están necesariamente dirigidos a uno mismo en tanto que respuestas a lo encomiables o indignos que somos. Esta idea implica que el orgullo o la vergüenza no son respuestas a lo que hacemos, *sino a nuestra propia imagen*. Emociones como el miedo, el enfado o la alegría, en cambio, harían referencia habitualmente a objetos diferentes del yo, pues en su caso, el objeto que ejemplifica la propiedad evaluativa relevante – ser peligroso, ofensivo o motivo de celebración – no es el sujeto. La revisión que llevo a cabo de los argumentos a favor de esta visión pretende mostrar que esta manera de entender la reflexividad de la vergüenza y el orgullo se enfrenta a serias dificultades. En lugar de explorar directamente su compromiso con la concepción escindida del sujeto, mi estrategia consiste en mostrar que tales argumentos no logran establecer que la vergüenza y el orgullo estén necesariamente dirigidos al yo como objeto intencional de la emoción (Tesis 1) y en señalar cuán problemático resulta trazar una línea que delimita con nitidez las características que le pertenecen propiamente al sujeto y las características que le son ajenas.

En segundo lugar, los participantes en el debate sobre las acciones expresivas sostienen que *el sujeto está desconectado de los gestos* a través de los cuales expresa emociones (**escisión sujeto-cuerpo**) e ignoran el papel que los demás – la *intersubjetividad* – desempeñan en el reconocimiento de tales gestos como respuestas de un sujeto (**escisión sujeto-otros**). Según la visión más extendida, no podemos concebir al sujeto como agente de sus sonrisas, llantos, risas o ceños fruncidos porque éstos son movimientos no instrumentales e involuntarios que sólo podemos explicar causalmente. Esto quiere decir que las expresiones faciales y corporales de las emociones son sucesos que meramente le acontecen al sujeto: están escindidos de su respuesta al mundo y de su perspectiva evaluativa sobre el mismo. En Cabrera (2022b), intento mostrar que, dados los presupuestos del debate sobre acciones expresivas, somos agentes, y no pacientes, de las expresiones faciales y corporales de las emociones (Tesis 2). La revisión que llevo a cabo de los presupuestos de este debate pone de manifiesto la tendencia generalizada a contemplar el cuerpo como algo desconectado de la vida mental del sujeto, lo cual parece explicar el tratamiento tan distinto que han recibido las expresiones emocionales más básicas, por una parte, y las acciones de tipo más sofisticado, por otra.

La exclusión de las sonrisas o los ceños fruncidos de la categoría de acciones, además, tiene que ver con el hecho de que, de manera estándar, se considera que el criterio para determinar si algo es una acción o un suceso es si el sujeto ejemplifica una determinada combinación de creencias y deseos. Trasladar el foco a las *prácticas* (esto es, a la *intersubjetividad*) en las que se inscriben las expresiones faciales y corporales de las emociones nos ayuda a ver lo problemático que resulta reducir tales expresiones a meros sucesos y a apreciar la relación constitutiva que existe entre el sujeto, su cuerpo y los demás.

En tercer lugar, el debate contemporáneo sobre las emociones concibe *a las propiedades evaluativas como independientes del sujeto y de sus respuestas emocionales* (**escisión sujeto-mundo**). De manera estándar, se asume que los distintos tipos de emoción se individúan a partir de las diferentes propiedades evaluativas – el miedo es una respuesta a lo peligroso; el enfado, a lo ofensivo; la humillación, a lo humillante, etc. – y que las propiedades evaluativas sobrevienen de propiedades naturales y se individúan independientemente de las respuestas emocionales del sujeto. Esta idea ha llevado a algunas de las teorías actuales de las emociones a proponer diferentes explicaciones de cómo las emociones registran o nos ponen en contacto con las propiedades evaluativas ejemplificadas por el objeto de la emoción. Como dicen Deonna y Teroni, “intuitively,

no other emotional experience than that of fear is a suitable candidate for presenting the world in terms of a danger” (2012: 86). Las teorías que asumen esta visión de las propiedades evaluativas, no obstante, tienden a estancarse en sus intentos por explicar cómo las emociones dan acceso a un tipo de propiedades totalmente independientes de ellas. En Cabrera (2022a) llevo a cabo una breve revisión de este compromiso filosófico así como de los argumentos que se han presentado en contra de la visión contraria, según la cual existe una relación de *dependencia mutua* entre respuestas emocionales y propiedades evaluativas. Dicha revisión no pretende establecer cuál de las dos concepciones de la relación entre emociones y propiedades evaluativas es preferible, sino mostrar que no es evidente que debamos rechazar la concepción de la dependencia mutua entre emociones y propiedades en favor de la postura dominante. Esta revisión sugiere, además, que los problemas a los que se enfrentan varias de las teorías actuales de las emociones podrían tener su origen, no tanto en la supuesta complejidad de las emociones, sino en su compromiso con la concepción de las propiedades evaluativas como independientes de las respuestas emocionales del sujeto (Tesis 3).

Como vemos, los tres debates analizados aquí se encuentran atravesados por una concepción escindida del sujeto. Esta manera de entender al sujeto por oposición al mundo, el cuerpo y los demás constituye una concepción problemática y empobrecida por diversos motivos. En primer lugar, cuando se opone el sujeto a los objetos de su vergüenza o de su orgullo, el sujeto acaba perdiendo parte del vínculo íntimo que le une a tales objetos y queda desproporcionalmente centrado en sí mismo. Este movimiento, además, lo desconecta de los demás en un sentido relevante. En segundo lugar, al separar al sujeto de su cuerpo y de los demás, el sujeto aparece como una pobre víctima de una serie de cambios faciales y corporales que nada tienen que ver con él. Y, en tercer lugar, cuando el sujeto queda escindido de las propiedades evaluativas, el sujeto queda reducido a un mero detector y desaparece su participación esencial en la creación de sentido y de valores. La importancia de revisar esta concepción del sujeto radica, además, en que se trata de una concepción que condiciona de manera significativa el estudio actual de las emociones, en concreto, al determinar el tipo de preguntas que guían la investigación. Como veremos con más detalle a continuación, en el debate sobre las emociones reflexivas, la escisión sujeto-mundo nos ha conducido a una caracterización de la vergüenza y el orgullo que otorga al yo un lugar muy implausible. En el debate sobre las acciones expresivas, por otra parte, las escisiones sujeto-cuerpo y sujeto-otro han permitido que se imponga una concepción muy intelectualista de nuestra agencia que no

es capaz de explicar el conjunto de cosas que hacemos como agentes. Y, por último, en el debate sobre la relación entre emociones y propiedades evaluativas, la escisión sujeto-mundo ha llevado a varias teorías recientes de las emociones a estancarse intentando explicar la manera en la que las emociones dan acceso a propiedades evaluativas.

Por estos motivos, la defensa que llevo a cabo de las tres tesis anteriores sugiere, además, que *una concepción más integrada o del sujeto* que tenga en cuenta su *relación constitutiva* con el mundo, su cuerpo y los demás sería preferible a la concepción escindida apenas descrita. Como es evidente, una concepción más integrada marcaría una dirección diferente tanto en los debates aquí analizados como en la reflexión filosófica sobre las emociones en general. Esta cuestión, no obstante, excede los límites de esta tesis y será un proyecto para otro momento. En lo que sigue, expongo con detalle el contenido y las conclusiones más importantes de cada uno de los artículos que componen esta tesis.

4. On the Intentionality of Shame and Pride

En el primero de los artículos (Cabrera en prensa), reviso uno de los presupuestos más extendidos en el debate actual sobre las emociones reflexivas. Según un importante número de autores, la diferencia fundamental entre la vergüenza y el orgullo, y el resto de nuestras emociones – miedo, odio, sorpresa, alegría, celos, enfado o compasión – es que las primeras se dirigen necesariamente al yo como objeto intencional de la emoción, mientras que las segundas se encuentran orientadas habitualmente hacia objetos del mundo distintos del yo (Helm 2001, Zinck 2008, Deonna y Teroni 2012, Deonna, Rodogno y Teroni 2012, Teroni 2016 y Tietjen 2020). En este artículo intento mostrar que los argumentos proporcionados a favor de esta idea no consiguen mostrar que la vergüenza y el orgullo se dirijan necesariamente a uno mismo – el agente (*the doer*) – y no a un objeto distinto del yo – el acto (*the deed*). Sugiero, además, que una explicación satisfactoria de tales emociones debería ser capaz de explicar los casos normativamente paradigmáticos en los que el sujeto de la vergüenza y el orgullo dirige su atención hacia lo que ha hecho y no hacia su propia imagen.

Para mostrar que la vergüenza y el orgullo no están necesariamente dirigidos al yo como objeto intencional de la emoción (Tesis 1), reviso tres argumentos a favor de la posición más extendida y ofrezco una serie de contraargumentos. Como veremos, los defensores de esta visión parecen, por una parte, otorgar al yo un lugar equivocado en la vergüenza y el orgullo y, por otra, basarse en un supuesto problemático según el cual las

características del objeto de una emoción se puede diferenciar nítidamente de las características del sujeto que la siente.

4.1. El argumento de la corrección

En primer lugar, según el Argumento de la corrección (*The Correctness Argument* o CA), la vergüenza y el orgullo están necesariamente dirigidos al sujeto de la emoción porque, para ser evaluadas como emociones correctas, el sujeto ha de ejemplificar la propiedad evaluativa que constituye el objeto formal de la emoción, a diferencia de lo que ocurre en el caso de las emociones no reflexivas (Teroni 2016, 5). Por ejemplo, la vergüenza que siento por la mentira que le he contado a una amiga es correcta si *yo* soy una persona indigna o vergonzosa (*unworthy* o *shameful*). En cambio, el miedo que siento ante el perro que tengo delante es correcto si *el perro* es peligroso. Como podemos ver, en el primer caso, la propiedad evaluativa – ser indigna o vergonzosa – es ejemplificada por el sujeto de la emoción mientras que, en el segundo, la propiedad evaluativa – ser peligroso – es ejemplificada por un objeto distinto del sujeto. De esto se sigue, según CA, que yo soy el objeto intencional de la vergüenza y que el perro es el objeto intencional de mi miedo. Este argumento parece asumir, por tanto, que el objeto que proporciona el criterio de corrección de una emoción – es decir, el objeto que ejemplifica la propiedad evaluativa en cuestión – *es el objeto* de la emoción.

Este argumento, sin embargo, ignora el hecho de que en el caso de la vergüenza y del orgullo siempre hay dos objetos que pueden ejemplificar la propiedad evaluativa relevante. Cuando siento vergüenza por la mentira que le he contado a mi amiga, no sólo yo sino *lo que he hecho* – mentir y causar un daño importante – constituye algo indigno o vergonzoso. Del mismo modo, cuando me enorgullezco de mi negocio, no sólo yo sino el propio negocio es algo encomiable (*praiseworthy*). Aquello que hago, sin embargo, parece tener una importante prioridad a la hora de determinar la corrección o proporcionalidad de mi respuesta emocional: el hecho de que contarle una mentira y herir a mi amiga sea indigno es, por una parte, lo que me convierte en una persona indigna (y no al revés) y, por otra, lo que determina que debería sentir vergüenza ante lo que he hecho. Esto quiere decir que, a pesar de que en el caso de la vergüenza o el orgullo el sujeto ejemplifique la propiedad evaluativa en cuestión, *ello no implica que esté respondiendo emocionalmente a sí mismo*.

Este razonamiento, más que determinar de manera definitiva cuál es el objeto de la vergüenza y del orgullo – yo misma o lo que he hecho –, muestra la dificultad de

distinguir claramente entre las características del sujeto y las características del objeto de su emoción. Las propiedades evaluativas que determinan la corrección de las respuestas emocionales parecen depender tanto del sujeto como del objeto al que responde emocionalmente, por lo que cualquier concepción que insista en que las propiedades evaluativas han de ser ejemplificadas, o bien por el sujeto, o bien por el objeto de la emoción, no podrá ofrecernos un análisis muy plausible de estas emociones.

4.2. El argumento fenomenológico

En segundo lugar, el Argumento fenomenológico (*The Phenomenological Argument* o PA) afirma que el hecho de que la vergüenza y el orgullo están necesariamente dirigidos al yo se manifiesta en la manera en la que dirigimos nuestra atención hacia nosotros mismos cuando sentimos tales emociones, a diferencia de lo que ocurre cuando sentimos miedo, alegría, compasión, etc. (Teroni 2016, 6). Es decir, dado que cuando siento vergüenza mis pensamientos y mi atención giran en torno a mí misma – “¿qué van a pensar ahora de mí?”, “¿cómo he podido hacer esto?”, “¡tendría que haberlo pensado mejor!” –, parece evidente que mi vergüenza *versa sobre mí misma*, es decir, está intencionalmente orientada hacia mí.

Este argumento, no obstante, ignora los casos de vergüenza y orgullo en los que el sujeto se centra en aquello que ha hecho en lugar de volverse narcisista o temerosamente hacia sí mismo y su propia imagen (Moran 2001, Corbí 2012). Parece que hay una diferencia relevante entre centrar mi atención en mí misma al sentir vergüenza por la mentira que le he contado a mi amiga – en el tipo de persona que soy, en lo que todo el mundo pensará de mí, etc. – y centrar mi atención en mi amiga y el daño que le he causado – en aquello que necesita de mí ahora, en la manera de reparar el daño, etc. Normalmente, censuramos el primer tipo de casos por considerarlos respuestas inapropiadas o desproporcionadas a la situación en la que nos encontramos: aunque solemos preocuparnos por nuestra imagen, ésta no parece ser la manera en la que *debamos* responder cuando hemos causado un daño a alguien. Esto quiere decir que los ejemplos de vergüenza en los que el sujeto se centra en el daño causado son casos *normativamente paradigmáticos* que nos permiten criticar los casos en los que el sujeto se centra en sí mismo en tanto que desviaciones. El hecho de que exista la tendencia psicológica a responder de manera desviada cuando sentimos vergüenza u orgullo, sin embargo, *no implica que el sujeto sea el objeto intencional* de las emociones en cuestión.

4.3. El argumento grammatical

Por último, según el Argumento grammatical (*The Grammatical Argument* o GA), la vergüenza y el orgullo están necesariamente dirigidos a uno mismo porque siempre podemos introducir un elemento reflexivo explícito en las frases que hacen referencia a estas emociones, a diferencia de lo que ocurre con las emociones no-reflexivas (Teroni 2016, 5). La idea es que podemos transformar “María está orgullosa del cuadro” en “María está orgullosa *de sí misma* por el cuadro que ha pintado” y “Marcos se avergüenza de su nariz grande” en “Marcos se avergüenza de sí mismo por su nariz grande” sin cambios relevantes en el significado de estas frases.

Respecto a la primera frase, no obstante, parece evidente que su significado cambia de manera importante tras la transformación: al principio, el orgullo de María parece estar intencionalmente dirigido al cuadro que ha pintado mientras que, tras la transformación, su emoción parece estar dirigida a alguna de sus habilidades o rasgos: ser buena pintora, haber aprovechado las clases recibidas, etc. Podemos preguntarnos, además, si sentir una emoción hacia alguno de nuestros rasgos – nuestras habilidades pictóricas, el tamaño de nuestra nariz, etc. – equivale a sentir una emoción hacia nosotros mismos, tal y como presupone GA. Parece, no obstante, que avergonzarme de mí misma es algo mucho más profundo que avergonzarme de alguno de mis rasgos: mientras que la vergüenza que puedo sentir por cierta parte de mi cuerpo puede afectar de manera más o menos superficial a mi comportamiento y mis preocupaciones, valorarme globalmente como una persona vergonzosa o indigna tiene, claramente, un impacto mucho más significativo en mi vida.

Esto quiere decir que no parece que podamos llevar a cabo las transformaciones propuestas por GA sin que se produzcan importantes cambios en el objeto intencional de la vergüenza y el orgullo. Por otra parte, y del mismo modo que nuestra revisión de CA, el análisis de este argumento parece poner de manifiesto cuán problemático resulta tratar de trazar una línea que delimita con nitidez las características que le pertenecen propiamente al sujeto y las características que le son ajenas.

4.4. Conclusiones

A través de esta breve revisión he intentado mostrar que los argumentos propuestos no logran justificar la idea de que la vergüenza y el orgullo están siempre intencionalmente dirigidos al yo, mientras que el resto de nuestras emociones se orientan hacia objetos del

mundo distintos del yo. Parece, por tanto, que cualquiera que quiera insistir en esta concepción de la reflexividad de la vergüenza y el orgullo habría de aportar nuevos argumentos a su favor. De esta discusión se desprende, además, que una teoría satisfactoria sobre la reflexividad de estas emociones habría de ser capaz de explicar los casos que hemos llamado normativamente paradigmáticos. De otro modo, continuaremos aproximándonos al yo en las emociones tomando exclusivamente como referencia una serie de casos que no nos permiten estudiar las emociones de la vergüenza y el orgullo en toda su complejidad.

Como se desprende de este análisis, la reflexión actual en torno a la reflexividad de la vergüenza y del orgullo se encuentra atravesada por la escisión sujeto-mundo, la cual presupone que es posible diferenciar con nitidez las propiedades del sujeto de las propiedades de los objetos a los que responde emocionalmente. Esta escisión parece generar importantes problemas pues, por una parte, ha dado lugar a una visión muy implausible de la intencionalidad de la vergüenza o el orgullo: tal y como reflejan los casos paradigmáticos que hemos tratado, el sujeto no parece estar respondiendo siempre emocionalmente a sí mismo cuando siente tales emociones (y, cuando se centra en sí mismo, censuramos su respuesta en tanto que inapropiada). Por otra parte, dicha escisión nos obliga a centrar la discusión en cuestiones tan artificiales como si la propiedad evaluativa relevante es ejemplificada por el sujeto (el agente) o por el objeto de su emoción (su acto). Tal vez, la propiedad evaluativa de ser indigno o vergonzoso se aplica, más bien, a un conjunto complejo del que forman parte tanto el agente como el acto, y no resulta posible determinar un criterio que permita distinguir entre los distintos elementos para así fijar cuál de ellos ejemplifica realmente o prioritariamente la propiedad evaluativa en cuestión. En cualquier caso, y más allá de estas últimas especulaciones, parece razonable concluir que la revisión del debate sobre las emociones reflexivas nos muestra la necesidad de encontrar una alternativa a la concepción escindida del sujeto presupuesta en este debate.

5. Duchenne Smiles are Actions not Mere Happenings: Lessons from the Debate on Expressive Action

En el segundo artículo (Cabrera 2022b), sostengo que, a diferencia de lo que suele asumirse en el debate sobre las acciones expresivas, somos agentes, y no pacientes, de las expresiones faciales y corporales de las emociones (Tesis 2). Para mostrar esto, comparo las expresiones faciales y corporales de las emociones – sonreír, fruncir el ceño, llorar,

etc. – con acciones expresivas simples – saltar de alegría, cubrirse avergonzado la cara, rascarse la cabeza con frustración, etc. – y argumento que, dados los presupuestos del debate sobre las acciones expresivas, ambos tipos de comportamiento han de ser clasificados de la misma manera, esto es, como acciones o como sucesos. Esto último, sin embargo, contradice la manera en la que nos relacionamos con conductas como saltar de alegría o cubrirse avergonzado la cara, por lo que concluyo que tanto las expresiones faciales y corporales de las emociones como las acciones expresivas simples han de ser incluidas en la categoría de acciones.

5.1. Expresiones faciales y corporales de las emociones

En el seno del debate sobre las acciones expresivas, existe un cierto consenso respecto a la idea de que las expresiones faciales y corporales de las emociones – sonreír, fruncir el ceño, llorar, apretar los puños, agachar la cabeza, etc. – no son acciones de un agente sino meros sucesos que le acontecen (Hursthouse 1991, Goldie 2000b, Betzler 2007 y 2009, Helm 2001 y 2016, Bennett 2016, Müller y Wong en prensa). Los criterios que se han proporcionado para justificar la exclusión de este tipo de comportamientos de la categoría de acciones son los siguientes: las expresiones faciales y corporales de las emociones (C1) no son conductas instrumentales, (C2) son ajenas a la racionalidad y (C3) son involuntarias. Según esto, las sonrisas o los ceños fruncidos son análogos a los espasmos y a los reflejos en tanto que meros sucesos que están desconectados de nuestra agencia. Desde esta perspectiva, las sonrisas no son algo que *hagamos* sino algo que *nos pasa*.

Esta visión tan extendida, sin embargo, genera un dilema importante en el debate de las acciones expresivas, pues no se puede afirmar que las expresiones faciales y corporales de las emociones no son acciones mientras se sostiene que las acciones expresivas simples sí lo son, dado que, como argumentaremos a continuación:

- Ninguna puede ser explicada instrumentalmente (ambas satisfacen C1)
- Ambas pueden ser explicadas racionalmente (ninguna satisface C2)
- Ambas pueden involucrar músculos que no podemos controlar directamente (ambas pueden satisfacer C3)

El hecho de que ambos tipos de comportamiento satisfagan exactamente los mismos criterios implica que *o ambos son acciones o ninguno lo es*. Esto último, sin embargo, resulta contraintuitivo, dado que tendríamos que asimilar acciones expresivas simples como saltar de alegría o cubrirse avergonzado la cara a meros sucesos, lo cual contradice

la manera en la que nos relacionamos con este tipo de comportamientos. Por este motivo, concluiremos que tanto las expresiones faciales y corporales de las emociones como las acciones expresivas simples han de ser incluidas en la categoría de acciones. Antes de poder alcanzar esta conclusión, sin embargo, analizamos brevemente en qué consisten las acciones expresivas.

5.2. Acciones expresivas simples

Tal y como ha señalado Hursthouse (1991), acciones como saltar de alegría o cubrirse avergonzado la cara plantean un importante problema a la concepción estándar de la acción (*the standard account of agency*). Según esta concepción, una conducta sólo puede ser considerada una acción si puede ser racionalizada en términos de un deseo y una creencia instrumental (Davidson 1963). Por ejemplo, cuando salgo corriendo ante la presencia de un depredador, mi comportamiento constituye una acción – y no un mero suceso – porque podemos racionalizarlo a partir de mi deseo de protegerme y mi creencia de que la mejor manera de protegerme es salir corriendo. Saltar de alegría o cubrirse avergonzado la cara, sin embargo, no pueden ser racionalizados de esta manera, pues no hay nada que el agente persiga mediante su conducta. Estos comportamientos, no obstante, son claramente acciones, dice Hursthouse. En concreto, y dado que no pueden ser racionalizados instrumentalmente, son acciones arracionales: acciones ajena a la racionalidad.

Esta propuesta generó, sin embargo, un enorme debate pues, desde el punto de vista de un importante número de autores, la idea de que las acciones expresivas están desconectadas de la racionalidad no resulta convincente. Cuando vemos que alguien salta de alegría al recibir buenas noticias vemos que lo que hace *tiene sentido* dada la emoción que siente, es decir, su conducta nos resulta inteligible. Esta idea ha llevado a muchos autores a proponer diferentes formas en las que podemos racionalizar este tipo de comportamientos y preservar, al mismo tiempo, la concepción estándar de la acción (Smith 1998, Goldie 2000, Helm 2001 y 2016, Betzler 2007 y 2009, Döring 2003, Scarantino y Nielsen 2015, Bennett 2016 y 2021, Pineda 2019, Müller y Wong en prensa).

En primer lugar, autores como Smith (1998) o Bennett (2016) han intentado proporcionar explicaciones instrumentales de las acciones expresivas a fin de mostrar que pueden ser racionalizadas en términos de la concepción estándar. Este tipo de explicación, sin embargo, conduce irremediablemente a la sobreintelectualización de las acciones expresivas simples. Betzler (2007 y 2009) y Goldie (2000), por otra parte, han sugerido

que, tal vez, saltar de alegría, rascarse la cabeza con frustración o cubrirse avergonzado la cara – dado que no exhiben la racionalidad propia de acciones más complejas – no sean acciones sino sucesos. Esta propuesta, no obstante, parece tener consecuencias bastante implausibles: si saltar de alegría no es más que un mero suceso, necesitamos una explicación de la manera en la que normalmente nos aproximamos a dicho comportamiento. Es decir, necesitamos una explicación de por qué, cuando vemos a alguien saltar de alegría, nos relacionamos con su comportamiento como una respuesta a algo que el sujeto valora positivamente y que puede ser evaluada como apropiada o inapropiada, a diferencia de lo que ocurre cuando vemos a alguien sufrir un espasmo.

Este tipo de problemas han llevado a Helm (2001 y 2016) y Müller y Wong (en prensa) a cuestionar la idea de que, para que un comportamiento sea considerado racional, éste haya de ser racional desde un punto de vista *instrumental*.

- *Supuesto de la racionalidad no-instrumental (Non-instrumental Rationality Assumption)*: saltar de alegría – al igual que el resto de nuestras acciones expresivas simples – puede considerarse una acción porque tiene un sentido (*a point*) que muestra que el agente está respondiendo a razones a través de su conducta y que ésta no es accidental

Gracias a este supuesto, parece que podemos preservar la intuición de que las acciones expresivas simples son acciones y evitar su sobreintelectualización.

5.3. Expresiones faciales y corporales de las emociones vs. acciones expresivas simples

Una vez concluido nuestro análisis de las acciones expresivas simples, podemos pasar a argumentar – a través de los argumentos A1-A3 – que, en virtud de C1-C3, tanto las acciones expresivas simples como las expresiones faciales y corporales de las emociones han de ser incluidas en la categoría de acciones.

(A1) Si aceptamos que las acciones expresivas simples son acciones a pesar de no ser instrumentales, no podemos excluir a las expresiones faciales y corporales de las emociones de la categoría de acciones por este motivo (como pretende C1).

Como acabamos de ver, si asumimos el supuesto de la racionalidad no-instrumental, las acciones expresivas simples pueden ser consideradas acciones a pesar de que, a través de ellas, el agente no pretenda satisfacer ningún objetivo. Esto quiere decir que no podemos

excluir a las expresiones faciales y corporales de las emociones de la categoría de acciones simplemente porque carecen de una estructura instrumental.

(A2) Si una acción puede ser racional simplemente por tener un sentido (a point) sin ser instrumentalmente racional, no tenemos razones (contra C2) para, en principio, excluir a las expresiones faciales y corporales de las emociones de la categoría de acciones.

Aunque, de manera unánime, los participantes en el debate sobre las acciones expresivas consideran que las expresiones faciales y corporales de las emociones son ajenas a la racionalidad (Helm 2001 y 2016, Bennett 2016 y Müller and Wong en prensa), si prestamos atención a las prácticas en las que se inscriben tales expresiones podemos ver cómo esto no puede ser correcto: (a) nos relacionamos con las expresiones faciales y corporales de las emociones como *respuestas* a algo y (b) las consideramos *reveladoras de la perspectiva evaluativa del sujeto* sobre la situación en la que se encuentra. Además, (c) parecen estar reguladas por una serie de *consideraciones morales* y (d) pueden ser *educadas*. Si las expresiones faciales y corporales de las emociones fueran realmente meros sucesos, las prácticas apenas descritas resultarían ininteligibles.

(A3) Si las acciones deben ser voluntarias, entonces las expresiones faciales y corporales de las emociones no son acciones, pero tampoco lo son muchas conductas instrumentales que, según la concepción estándar de la acción, deberían ser consideradas acciones (contra C3).

Como Goldie indica, cuando sonreímos de manera genuina, no podemos controlar directamente los músculos involucrados en nuestra sonrisa (2000b, 35). Esto muestra, según él, que las sonrisas no son acciones sino meros sucesos. No obstante, hay conductas instrumentales – las cuales, desde la perspectiva de la concepción estándar de la acción, son claramente acciones – en las que no podemos controlar tampoco de manera directa la musculatura de nuestro cuerpo: cuando salgo corriendo ante la amenaza de un depredador, debido al aumento de adrenalina, el modo en el que muevo mi cuerpo es diferente al modo en que puedo moverlo cuando, por ejemplo, salgo a hacer footing. Este hecho, sin embargo, no nos impide considerar que lo que hago – huir del depredador – es una acción y no un mero suceso. Por tanto, no parece correcto excluir a las expresiones faciales y corporales de las emociones de la categoría de acciones simplemente porque no podemos controlar directamente todos los músculos que intervienen en el movimiento.

5.4. Conclusiones

A través de los argumentos A1-A3, he intentado mostrar que, en virtud de los criterios C1-C3, las expresiones faciales y corporales de las emociones no pueden ser clasificadas de manera distinta a las acciones expresivas simples. Dado que asimilar estas últimas a meros sucesos resulta altamente problemático, he concluido que ambos tipos de comportamiento deberían ser considerados acciones. De esto se sigue que somos agentes, y no pacientes, de las expresiones faciales y corporales de las emociones.

La tendencia generalizada a tratar las expresiones faciales y corporales de las emociones de manera diferente del resto de nuestras acciones revela la manera profunda en la que la escisión sujeto-cuerpo condiciona el debate sobre las acciones expresivas. La idea de que las sonrisas, los llantos o los ceños fruncidos no son expresiones de nuestra agencia presupone una visión del cuerpo como algo separado del sujeto – como algo que, como mucho, revela a través de una serie de síntomas la existencia de un sujeto pero que no se identifica con él. Como hemos podido ver, esta manera de concebir las expresiones faciales y corporales de las emociones resulta muy implausible incluso desde los propios presupuestos del debate sobre las acciones expresivas.

Por otra parte, tanto los participantes en este debate como la concepción estándar de la acción se aproximan al sujeto como una realidad escindida de los demás y examinan la cuestión de su agencia apelando exclusivamente a la exemplificación de una serie de estados mentales.³ Cuando observamos al sujeto en su relación con los demás, no obstante, se vuelve evidente que sus sonrisas, risas y llantos no pueden ser meros sucesos que le acontecen: las prácticas que regulan tales expresiones nos muestran que se tratan de *respuestas* de un sujeto y que, por tanto, la visión más extendida en el debate sobre las acciones expresivas es claramente errónea. La conclusión más importante de esta revisión es que parece que una buena explicación de las expresiones faciales y corporales de emoción deberá aproximarse al sujeto desde la relación constitutiva que éste mantiene con su cuerpo y los demás.

³ Con la notable excepción de Helm (2001 y 2006) y Müller y Wong (en prensa).

6. ¿Qué es una emoción?

El objetivo del tercer artículo (Cabrera 2022a) consiste, por una parte, en presentar un estado de la cuestión sobre la investigación filosófica en emociones⁴ y, por otra, en llevar a cabo una breve revisión de los compromisos filosóficos que pueden estar condicionando algunas de las teorías de las emociones más importantes. En particular, examino el compromiso generalizado con una concepción de las propiedades evaluativas como independientes de las respuestas emocionales del sujeto y argumento que los problemas a los que se enfrentan varias de las teorías actuales de las emociones podrían tener su origen en dicha concepción y no en la supuesta complejidad de las emociones (Tesis 3). Comenzamos con unas breves aclaraciones iniciales y examinamos, a continuación, las diferentes teorías.

6.1. Aclaraciones iniciales y requisitos para una buena teoría

En primer lugar, este artículo se aproxima al estudio de las emociones a través de un tipo de explicación personal. Para poder dar cuenta del discurso normativo sobre emociones – “no deberías sentirte culpable por esto”, “el comentario que le hizo fue muy ofensivo, normal que se enfadara”, etc. –, nuestras explicaciones han de tener en cuenta que el sujeto tiene un determinado *punto de vista* sobre la situación en la que se encuentra y que es él, y no meramente su cuerpo o una serie de procesos causales ajenos a su perspectiva evaluativa, quien responde emocionalmente. A pesar de que el conjunto de las teorías que examinamos en este artículo pretenden dar cuenta de este tipo de discurso, veremos cómo algunas de ellas no parecen cumplir realmente este objetivo. En segundo lugar, reflexionamos en torno a las emociones entendidas como episodios y no como disposiciones. Aunque no está claro que lo que las emociones son se agote en los llamados episodios emocionales (Wollheim 1999, Goldie 2000a), la discusión contemporánea sobre emociones está organizada en torno a la idea de que las emociones son estados episódicos de corta duración que se sienten de una determinada manera.

Otra de las ideas más extendidas en el debate actual sobre emociones es que cualquier teoría satisfactoria ha de ser capaz de explicar tres aspectos fundamentales de

⁴ A diferencia de lo que ocurre en el ámbito anglosajón, en el que abundan las entradas en enciclopedias y manuales sobre filosofía de las emociones, existen pocos textos de referencia en castellano que proporcionen una visión general del estado de la investigación filosófica sobre las emociones. Los trabajos de Vendrell (2018) y Pineda (2019) son una muestra del creciente interés por hacer accesible esta disciplina a los lectores en castellano.

las emociones. El primero de ellos es su aspecto intencional y hace referencia al hecho de que las emociones – del mismo modo que los recuerdos, las percepciones o las creencias – están dirigidas o son respuestas a objetos, situaciones, eventos o estados de cosas. El miedo, por ejemplo, es siempre miedo *de algo* y no un mero sentir bruto. En segundo lugar, las emociones son fenómenos evaluativos, esto es, siempre representan el objeto al que están dirigidas bajo un determinado aspecto evaluativo (Kenny 1963). En este sentido, el miedo es una respuesta a lo peligroso; la humillación, a lo humillante; la alegría, a lo que es motivo de celebración, etc. Esto nos permite individuar los distintos tipo de emoción en tanto que respuestas a diferentes propiedades evaluativas. Por último, el aspecto sentido o fenomenológico de las emociones hace referencia al hecho de que las emociones son *experiencias* con una cierta cualidad sentida. Hay quien intenta explicar dicha cualidad exclusivamente en términos de sensaciones de cambios corporales (James 1884, Deonna y Teroni 2012) y quien intenta describir sentires de tipo más “vivencial” que no refieran únicamente a cómo se siente el cuerpo (Goldie 2000a, Deigh 2010, Vendrell 2018).

6.2. La teoría creencia-deseo

La teoría creencia-deseo identifica las emociones con combinaciones de creencias y deseos: el miedo a un perro consistiría en la creencia de que el perro va a abalanzarse sobre nosotros y el deseo de salir corriendo, y el arrepentimiento por haber mentido a una amiga, en la creencia de que no hemos dicho la verdad y el anhelo de haber actuado de manera diferente (Gordon 1987, Green 1992. Ver también Davidson 1976). Según esta teoría, los cambios corporales que acompañan a las emociones son meros efectos causales de estas combinaciones de creencias y deseos. El aspecto sentido de las emociones, desde esta perspectiva, se reduce a la percepción de tales cambios.

Esta manera de entender las emociones, no obstante, ha sido ampliamente rechazada por diversos motivos: una fenomenología de tipo corporal no parece agotar la manera en la que se sienten las emociones, las combinaciones de creencias y deseos no son necesarias ni suficientes para tener emociones, ni nos permiten distinguir entre los distintos tipos de emoción y la posibilidad de sentir emociones queda reducida a la posesión de lenguaje, pues la capacidad para tener creencias y deseos presupone un pensamiento de tipo proposicional. Esto último implicaría que, según los presupuestos de la teoría creencia-deseo, los animales y los niños de corta edad no pueden sentir emociones, lo cual es claramente incorrecto (Deigh 2010). Aunque esta última objeción

suele presentarse como una dificultad insuperable para la teoría creencia-deseo, el presupuesto fundamental en el que esta se basa – que la capacidad para tener creencias y deseos depende de la posesión del lenguaje – es un presupuesto filosóficamente controvertido (Dennett 1995).

6.3. La teoría juicialista

La teoría juicialista, por otra parte, sostiene que las emociones son un tipo de juicio evaluativo sobre objetos y situaciones que nos conciernen: el miedo al perro consistiría en juzgar que el perro es peligroso y el arrepentimiento por haber mentido a una amiga consistiría en juzgar que la mentira es algo lamentable (*regretful*) (Solomon 1993, Nussbaum 2001). Al incluir conceptos evaluativos en los juicios que constituyen las emociones, esta teoría nos permite individuar los distintos tipos de emoción (Deigh 1994, Roberts 2003). No obstante, y del mismo modo que la teoría anterior, el juicialismo considera que los cambios corporales que acompañan a las emociones son efectos causales – y no parte constitutiva – de las mismas, y que el aspecto sentido de las emociones se reduce a la percepción de estos cambios. Este es uno de los motivos por los que la teoría juicialista ha recibido numerosas objeciones.

Los críticos de esta teoría han señalado, además, que los juicios evaluativos no parecen ser necesarios ni suficientes para experimentar emociones, que la explicación de las emociones recalcitrantes que se deriva del juicialismo no resulta satisfactoria y que la capacidad para tener emociones queda reducida a la posesión de conceptos evaluativos y, por tanto, de lenguaje. Esta última crítica, sin embargo, y como ya hemos visto, se basa en un presupuesto filosóficamente controvertido. En sus intentos por matizar el tipo de juicio que constituyen las emociones (Nussbaum 2001, Solomon 2004), los defensores de la teoría juicialista han sido acusados asimismo de introducir un tipo de juicio *ad hoc* para evitar las objeciones planteadas (Scarantino 2010). Este tipo de acusación, sin embargo, parece presuponer una noción de juicio muy concreta como paradigma. Aunque se suele acusar al juicialismo de sobreintelectualizar a las emociones, parece que los propios críticos de este teoría se aferran a una visión sobreintelectualizada de los juicios. Tal vez, la idea de que las emociones son un tipo de juicio nos permitiría modificar nuestro concepto de juicio si descubrimos que se trata de un concepto que sólo permite dar cuenta de una serie de casos muy intelectualizados.

6.4. La teoría perceptiva

La teoría perceptiva, en cambio, identifica las emociones con percepciones de valores o propiedades evaluativas (De Sousa 1987, Roberts 2003, Tappolet 2016). Según esta teoría, del mismo modo que percibimos el azul o la redondez de una pelota, percibimos la peligrosidad de un perro o lo lamentable que es mentir a una amiga. Los defensores de la teoría perceptiva sostienen que la percepción proporciona un buen modelo para entender las emociones porque ambos tipos de estados son intencionales, no requieren el uso de conceptos y su cualidad sentida queda inexplicada si son reducidos a meros juicios. Aunque esta teoría parecía una prometedora alternativa a la teoría juicialista, los presupuestos tan básicos en los que se basa constituyen una visión muy controvertida en filosofía de la percepción. En concreto, no está claro cómo la percepción entendida como un registro pasivo y no conceptual de lo que nos rodea podría justificar las creencias y los juicios que nos formamos a partir de nuestras experiencias perceptivas (McDowell 1994).

Tanto la teoría perceptiva como sus objetores, no obstante, presuponen este modelo de la percepción. Esto quiere decir que las críticas a esta teoría resultarán convincentes en la medida en la que simpaticemos o no con la visión de la percepción presupuesta en esta discusión. Tales críticas se esfuerzan por mostrar que hay diferencias profundas entre emociones y percepciones: las emociones requieren bases intencionales para acceder a su objeto mientras que las percepciones acceden directamente a su objeto, las emociones se caracterizan por su fenomenología “de respuesta” y las percepciones por su fenomenología “presentacional” y las emociones, a diferencia de las percepciones, no pueden estar dirigidas a propiedades sino que se dirigen exclusivamente a los portadores de las propiedades (Deonna y Teroni 2012, Mitchell 2021).

6.5. La teoría jamesiana

Según la teoría jamesiana, las emociones son percepciones de cambios corporales: el miedo al perro consiste en sentir cómo se nos acelera el corazón, se nos paraliza el cuerpo, se nos eriza el vello, etc. mientras que el arrepentimiento por haber contado una mentira consistiría, por ejemplo, en notar cómo nuestra cabeza se agacha, se nos encoge el estómago, etc. (James 1884). A pesar de que, a diferencia de la teoría creencia-deseo y la teoría juicialista, la teoría jamesiana otorga un lugar principal al aspecto sentido de las emociones, esta última se enfrenta a dos objeciones fundamentales: la primera es que parece que hay personas que, debido a una lesión de la médula espinal, no pueden sentir

sus cuerpos y, sin embargo, tienen emociones (Cannon 1927, Cobos et al. 2002) y la segunda es que esta teoría atribuye una intencionalidad errónea a las emociones (Pineda 2019). Si las emociones son percepciones de cambios corporales, entonces tales cambios constituyen el objeto intencional de las emociones, sin embargo, el discurso normativo que regula nuestras emociones muestra que éstas no son respuestas a cambios que tienen lugar en nuestro cuerpo sino a objetos, personas, eventos, etc.

La teoría neo-jamesiana de Prinz (2004) intenta evitar la segunda objeción adoptando una concepción teleosemántica de la intencionalidad según la cual los cambios corporales que acompañan a las emociones representan propiedades evaluativas. La idea es que los cambios corporales del miedo – temblor, sudoración, etc. – habrían sido seleccionados evolutivamente como indicadores de la presencia del peligro y, por tanto, tienen la función de llevar esta información. No obstante, en esta forma de representación no parece haber lugar para un sujeto que aprehende evaluativamente, desde una perspectiva de primera persona, el objeto de su emoción (Deigh 2010). Además, se pierde la inteligibilidad primitiva que parece existir entre la fenomenología de las emociones y las propiedades evaluativas correspondientes: la manera en la que se siente el cuerpo no parece ser un mero síntoma de la presencia de una propiedad evaluativa (Deonna y Teroni 2012).

6.6. La teoría actitudinal

La teoría actitudinal, por su parte, considera que las emociones son actitudes evaluativas corporalmente sentidas que no pueden reducirse a otro tipo de estado mental (Deonna y Teroni 2012, 2020). En concreto, las emociones son posicionamientos corporales – o tendencias a la acción (Frijda 2007) – orientados hacia el objeto de la emoción. El miedo, por ejemplo, consistiría en sentir el cuerpo prepararse para disminuir el impacto del perro que va a atacarnos. La vergüenza, en cambio, consistiría en sentir el cuerpo prepararse para que la tierra lo trague o para que desaparezca de la vista de los demás⁵. Según esta teoría, evaluamos el objeto de la emoción a través de la manera en la que nuestro cuerpo se prepara para actuar. La teoría actitudinal, sin embargo, es vulnerable a la objeción de las personas con una lesión de médula espinal que, a pesar de no sentir sus cuerpos, siguen teniendo emociones (Cannon 1927, Cobos et al. 2002), tampoco consigue explicar la

⁵ Aunque hasta ahora hemos hecho referencia al arrepentimiento, y no a la vergüenza, en nuestros ejemplos, Deonna y Teroni no explican en qué podría consistir el posicionamiento corporal correspondiente al arrepentimiento. He preferido poner un ejemplo de los propios autores de la teoría a fin de no distorsionarla.

correspondencia entre patrones de comportamiento y propiedades evaluativas, sostiene que es el cuerpo – y no el sujeto – quien se prepara para actuar e ignora el hecho de que nuestro discurso normativo sobre emociones no regula ni versa sobre posicionamientos corporales.

Todos estos problemas parecen estar relacionados con la reducción de la cualidad sentida de las emociones a una fenomenología de tipo corporal. Ante este tipo de dificultades, los defensores de otra familia de teorías – los teóricos de la intencionalidad afectiva – defienden que hemos de que hacer referencia a otro tipo de sentires que no sean meramente corporales para entender la manera en la que las emociones se dirigen a su objeto y nos ponen en contacto con sus propiedades evaluativas. Por ejemplo, a través de la aprobación o desaprobación sentida (Mitchell 2021) o la sensación de placer o dolor asociada a la experiencia emocional (Helm 2002) podríamos acceder a cómo es evaluativamente el objeto de nuestras emociones. No obstante, no resulta nada claro cómo sensaciones afectivas tan básicas como aprobar o desaprobar pueden ponernos en contacto con el conjunto de propiedades evaluativas que se corresponden con los diferentes tipos de emoción.

6.7. El problema de la relación entre emociones y propiedades evaluativas

La revisión de algunas de las teorías de las emociones más importantes muestra que, tanto las propias teorías, como algunas de las objeciones planteadas en su contra, asumen ciertos compromisos filosóficos que parecen condicionar de manera significativa el estudio de las emociones. En concreto, hemos señalado que la idea de que la posesión de lenguaje es necesaria para la atribución de creencias y deseos es filosóficamente controvertida, que los críticos de la teoría juicialista se basan en una concepción sobreintelectualizada de los juicios y que la visión de la percepción como un registro pasivo y no conceptual presupuesta por los defensores y los críticos de la teoría perceptiva es cuestionable. Más allá de estas cuestiones, sin embargo, parece haber un compromiso filosófico más profundo y complejo que condiciona seriamente las propuestas de la teoría perceptiva, de la teoría actitudinal y de las teorías de la intencionalidad afectiva.

Como se desprende de nuestra discusión, estas tres últimas teorías identifican la reflexión sobre el aspecto evaluativo de las emociones con la reflexión sobre la manera en la que las emociones *registrar o nos ponen en contacto con las distintas propiedades evaluativas* que nos permiten individuarlas. No obstante, todas estas teorías se atascan en este mismo punto. La teoría perceptiva, en primer lugar, no logra establecer una analogía

entre la manera en la que percibimos propiedades no evaluativas (azul) y propiedades evaluativas (peligrosidad). La teoría actitudinal, en segundo lugar, plantea una correspondencia muy implausible entre los diferentes posicionamientos corporales asociados a cada emoción y las correspondientes propiedades evaluativas. Por último, respecto a la teoría de la intencionalidad afectiva, no está claro cómo sensaciones afectivas tan básicas como aprobar o desaprobar, sentir placer o dolor, pueden ponernos en contacto con propiedades evaluativas complejas como ser peligroso o reprochable.

Este problema común a las diferentes teorías podría tener que ver con la complejidad de las emociones, sin embargo, se trata de un problema que sólo se plantea si uno está comprometido con *una concepción de las propiedades evaluativas como independientes de las respuestas emocionales del sujeto*. Según la visión dominante en el debate actual sobre emociones, las propiedades evaluativas son propiedades del mundo que nuestras emociones registran y que supervienen de propiedades naturales (Deonna y Teroni 2012, Tappolet 2016, Mitchell 2021): la propiedad de ser peligroso de un perro sobrevendría de la propiedad natural de tener colmillos afilados o un gruñido fuerte, mientras que la propiedad de ser humillante sobrevendría de la propiedad natural de que tu entorno te ignore o te llame por un nombre que no es el tuyo (Kim 1993).

Uno de los problemas fundamentales con esta concepción de las propiedades evaluativas es que parece que necesitamos una explicación del salto que parece haber entre lo natural y lo evaluativo (Smith 2014): el miedo no es una respuesta a lo que tiene la propiedad de morder, sino a lo que tiene la propiedad de ser peligroso y la humillación no es una respuesta a quien no te llama por tu nombre, sino a quien te humilla. Por muchas propiedades evaluativas que añadamos, no podremos explicar por qué cierta respuesta es un caso de miedo, de humillación o de cualquier otra emoción a menos que hagamos referencia a conceptos de tipo evaluativo (Berlin 1958, Corbí 2012).

Este tipo de problema ha dado lugar a una concepción alternativa de las propiedades evaluativas según la cual existe *una relación de dependencia mutua entre emociones y propiedades evaluativas* (Wiggins 1987, Goldie 2000a, Dunn 2006, Corbí 2012). La idea es que no podemos individuar las propiedades evaluativas sin hacer referencia a los distintos tipos de emoción (y no sólo al revés): lo ofensivo es lo que merece enfado, lo peligroso es lo que merece miedo, lo vergonzoso es lo que merece vergüenza, etc. y no lo que sobreviene de ciertas propiedades naturales. A pesar de que se trata de una concepción ampliamente rechazada en el debate actual sobre emociones, los argumentos que se han presentado en su contra no parecen ser definitivos o, al menos,

no parecen determinar que una concepción de las propiedades evaluativas como independientes de las emociones sea claramente preferible. La adopción de una u otra posición, sin embargo, condiciona significativamente el tipo de preguntas que guían la investigación sobre emociones, por lo que resulta pertinente revisar brevemente los argumentos ofrecidos en contra de la concepción de la dependencia mutua entre emociones y propiedades así como algunas de las respuestas ofrecidas.

(A1) La concepción de la dependencia mutua ofrece una explicación circular: algo tiene la propiedad de ser encomiable si merece orgullo, y algo merece orgullo si tiene la propiedad de ser encomiable.

En respuesta a este tipo de objeción, Wiggins señala que el tipo de circularidad involucrada en esta propuesta es una circularidad virtuosa (1987). Cuando tratamos de elucidar en qué consisten las propiedades evaluativas, hemos de observarlas en el contexto en el que tienen sentido. Esto quiere decir que las respuestas emocionales y las propiedades evaluativas han de ser observadas las unas a la luz de las otras: en el contexto apropiado, las preguntas “¿he de admirar x?” y “¿es x admirable?” son indistinguibles (Dunn 2006).

(A2) Si las propiedades evaluativas han de ser individuadas por las respuestas emocionales que producen, no puedo estar nunca equivocada con respecto a si algo tiene una determinada propiedad.

La concepción de la dependencia mutua entre emociones y propiedades evaluativas, sin embargo, no sostiene que estas últimas hayan de individuarse por las respuestas *efectivas* de cada sujeto. Según Wiggins, a través de un proceso de instrucción mutua y mejoramiento, aprendemos que los objetos y situaciones a los que respondemos con una determinada emoción poseen paradigmáticamente una determinada propiedad evaluativa. Una vez hemos aprendido esto, podemos comprobar si algo tiene una determinada propiedad fijándonos en las marcas propias de los casos paradigmáticos y no meramente en si un sujeto concreto experimenta o no una determinada emoción.

(A3) La idea de que las propiedades evaluativas dependen de las respuestas emocionales conduce irremediablemente al relativismo moral.

En respuesta a este argumento, conviene diferenciar claramente el relativismo moral del *pluralismo moral* (Berlin 1958, Corbí 2012): asumir que ante un determinado problema moral no existe una única respuesta válida, no implica que cualquier respuesta lo sea. Para

que una respuesta cuente como propiamente moral ha de estar vinculada a aspectos moralmente relevantes de la situación.

(A4) A menos que la concepción de la dependencia mutua aclare el tipo de razones que determinan qué emoción es la que merece una situación, dicha propuesta no es sostenible. Un demonio vil puede no ser admirable pero prudencialmente merecer admiración y una broma puede ser objetivamente divertida pero no merecer diversión dados nuestros estándares morales.

La cuestión importante aquí es si, en el primer caso, el defensor de la dependencia mutua entre emociones y propiedades evaluativas está obligado a concluir que el demonio *es admirable*. Sin embargo, y como hemos visto antes, parece que el defensor de esta postura puede reconocer situaciones en las que un sujeto particular responde con una determinada emoción sin que el objeto al que se dirige su emoción la merezca (Wiggins 1987). En el segundo caso, por otra parte, si todo lo que importara fuera no herir los sentimientos de otra persona, parece que estamos ante consideraciones de tipo prudencial, no moral: aunque sentir diversión sea correcto, no debería reírme delante de la persona a quien va dirigida la broma. No obstante, el motivo por el que resulta moralmente inapropiado vivir la broma como algo divertido parece más bien que ello me insensibiliza ante el daño que puede estar causado al otro y, en este sentido, podemos cuestionarnos en qué sentido relevante hemos de seguir insistiendo en que la diversión es objetivamente correcta aunque moralmente inapropiada

6.8. Conclusiones

Lejos de establecer cuál de las dos concepciones de la relación entre las emociones y las propiedades evaluativas es preferible, el objetivo de esta última revisión ha sido mostrar que estamos ante una cuestión abierta y que no es evidente que hayamos de rechazar la concepción de la dependencia mutua entre emociones y propiedades en favor de la postura dominante. Como hemos visto, sin embargo, varias de las teorías de las emociones más recientes están guiadas por la idea de que las propiedades evaluativas son independientes de las emociones y que parte de lo que tenemos que investigar cuando nos preguntamos qué son las emociones es de qué manera éstas registran o dan acceso a tales propiedades. Ante las dificultades que las teorías anteriores encuentran para dar cuenta de esto último, no podemos dejar de preguntarnos si esta dificultad es inherente a la complejidad del

fenómeno estudiado o si podría estar relacionada, más bien, con la concepción de las propiedades evaluativas presupuesta por tales teorías.

Como podemos ver, la concepción dominante de la relación entre emociones y propiedades evaluativas en el debate filosófico actual presupone una escisión de tipo sujeto-mundo pues, desde esta perspectiva, propiedades como ser humillante o digno de admiración son independientes de las respuestas emocionales de un sujeto. Esta manera de concebir al mundo en el que nos movemos, valoramos y nos relacionamos como independiente de nosotros supone, por una parte, un gran compromiso filosófico y, por otra, un compromiso que, tal y como hemos visto, no está libre de dificultades. El hecho de que varias de las teorías actuales de las emociones se atasquen en sus intentos por explicar la manera en la que las emociones dan acceso a propiedades evaluativas podría, sin embargo, indicar – tal y como sugieren los defensores de la dependencia mutua entre emociones y propiedades evaluativas – que una concepción del sujeto más integrado con su mundo permitiría alcanzar una mejor comprensión de sus emociones.

7. Conclusión

Las tesis que he defendido a través de estos tres artículos ponen de manifiesto cómo las diferentes maneras en las que se concreta la concepción escindida del sujeto condicionan de manera relevante la discusión contemporánea sobre emociones. Respecto al debate sobre las emociones reflexivas, hemos visto cómo la escisión sujeto-mundo ha conducido a una concepción muy implausible de la reflexividad de la vergüenza y el orgullo y a una distinción muy artificial entre sujeto y objeto de la emoción. En relación al debate sobre las acciones expresivas, las escisiones sujeto-cuerpo y sujeto-otro han contribuido a una visión de nuestras acciones que impide dar cuenta de todo lo que hacemos como agentes. Y, en el debate sobre la relación entre emociones y propiedades evaluativas, la escisión sujeto-mundo ha fomentado un tipo de investigación sobre las emociones que podría estar profundamente desencaminado.

Este tipo de consideraciones deberían, como mínimo, animarnos a revisar la concepción del sujeto con la que estamos implícitamente comprometidos cuando reflexionamos filosóficamente sobre las emociones. Estas consideraciones sugieren, además, la importancia de proporcionar una concepción más integrada del sujeto. Una concepción de este tipo habría de tener en cuenta la relación constitutiva o la profunda continuidad que parece existir entre el sujeto, el mundo, el cuerpo y los demás. Como

decíamos al comienzo de este resumen, una concepción más integrada plantearía una serie de preguntas y problemas diferentes a los planteados por la concepción escindida que hemos examinado aquí y marcaría, por tanto, una dirección diferente tanto en los debates que hemos analizado como en la reflexión filosófica sobre las emociones en general. Esta cuestión, no obstante, excede las limitaciones de esta tesis y será un proyecto para otro momento.

Referencias

- Bennett, C. (2016). “Expressive Actions”, en C. Abell & J. Smith (eds.) *Emotional Expression: Philosophical, Psychological and Legal Perspectives*, Cambridge University Press, p. 73-94.
- Bennett, C. (2021). The Problem of Expressive Action, *Philosophy*, 96(2), 277-300.
- Berlin, I. (1958). *The Crooked Timber of Humanity*, Princeton: Princeton University Press.
- Betzler, M. (2007). Making Sense of Actions Expressing Emotions, *Dialectica*, 61(3), 447-466.
- Betzler, M. (2009). Expressive Actions, *Inquiry*, 52(3), 272-292.
- Cabrera, M. (2022a). ¿Qué es una emoción?, *Quaderns de filosofia*, 8(2), 145-91.
- Cabrera, M. (2022b). Duchenne Smiles are Actions not Mere Happenings: Lessons from the Debate on Expressive Action, *Theoria: An International Journal for Theory, History and Foundations of Science*, 37(2), 163-179.
- Cabrera, M. (en prensa). On the Intentionality of Shame and Pride, *Daimon Revista Internacional de Filosofía*.
- Cannon, W. B. (1927). The James-Lang Theory of Emotion: A Critical Examination and an Alternative Theory, *American Journal of Psychology*, 39: 106-124.
- Cobos, P. et al. (2002). Revisiting the James versus Cannon Debate on Emotion: Startle and Autonomic Modulation in Patients with Spinal Cord Injuries, *Biological Psychology* 61(3): 251-69.
- Corbí, J. E. (2012). *Morality, Self-knowledge and Human Suffering: An Essay on the Loss of Confidence in the World*. Routledge.
- Davidson, D. (1963). Actions, Reasons and Causes, *Journal of Philosophy*, 60(23), 685-700.
- Davidson, D. (1976). Hume’s Cognitive Theory of Pride, *The Journal of Philosophy*, 73 (19): 744-757.

- De Sousa, R. (1987). *The Rationality of Emotion*, MIT Press.
- Deigh, J. (1994). Cognitivism in the Theory of Emotions, *Ethics*, 104: 824-54.
- Deigh, J. (2010). “Concepts of Emotions in Modern Philosophy and Psychology” en P. Goldie (ed.) *The Oxford Handbook of Philosophy of Emotion*, Oxford University Press, p. 17-40.
- Deonna, J. A., Rodogno, R. & Teroni, F. (2012). *In Defense of Shame: The Faces of an Emotion*, Oxford University Press.
- Deonna, J. A. y Teroni, F. (2012). *The Emotions: A Philosophical Introduction*, Routledge.
- Deonna, J. A. y Teroni, F. (2020). “Emotional Experience: Affective Consciousness And its role in Emotion Theory” en U. Kriegel (ed.) *Oxford Handbook of the Philosophy of Consciousness*, Oxford University Press.
- Dennett, D. C. (1995). “Do Animals Have Beliefs?”, en H. Roitblat (ed.) *Comparative Approaches to Cognitive Sciences*, MIT Press.
- Döring, S. (2003). Explaining Action by Emotion, *The Philosophical Quarterly*, 53(211), 214-230.
- Dunn, R. (2006). *Values and the Reflective Point of View: On Expressivism, Self-knowledge and Agency*, Routledge.
- Frijda, N. H. (2007). *The Laws of Emotion*, Lawrence Erlbaum Associates Publishers.
- Goldie, P. (2000a). *The Emotions*, Oxford: Oxford University Press.
- Goldie, P. (2000b). Explaining expressions of emotion, *Mind*, 109, 25-38.
- Gordon, R. (1987). *The Structure of Emotions: Investigations in Cognitive Philosophy*, Cambridge, Cambridge University Press.
- Green, O. H. (1992). *The Emotions: A Philosophical Theory*, Dordrecht, Kluwer.
- Helm, B. (2001). *Emotional Reason: Deliberation, Motivation, and the Nature of Value*, Cambridge University Press.
- Helm, B. (2002). Felt Evaluations: A Theory of Pleasure and Pain, *American Philosophical Quarterly*, 39: 13-30.
- Helm, B. (2016). Emotional Expression, Commitment and Joint Value, In C. Abell & J. Smith (Eds.). *Emotional Expression: Philosophical, Psychological and Legal Perspectives*, Cambridge University Press, p. 95-114.
- Hursthouse, R. (1991). Arational Actions, *Journal of Philosophy*, 88, 57-68.
- James, W. (1884), What is an Emotion, *Mind*, 9: 188-205.
- Kim, J. (1993). *Supervenience and Mind*, Cambridge University Press
- McDowell, J. (1994). *Mind and World*, Harvard University Press.

- Mitchel, J. (2021). *Emotion as Feeling Towards Value: A Theory of Emotional Experience*, Oxford University Press.
- Moran, R. (2001). *Authority and Estrangement: An Essay on Self-knowledge*, Princeton University Press.
- Müller, M. J. y Wong, H. Y. (forthcoming). “Which Emotional Behaviors are Actions?”, en A. Scarantino (ed.) *Routledge Handbook of Emotion Theory*, London/New York: Routledge.
- Nussbaum, M. C. (2001). *Upheavals of Thought: The Intelligence of Emotions*, New York, Cambridge University Press.
- Perry, J. (1979). The Problem of the Essential Indexical, *Nous* 13, 3-21.
- Prinz, J., (2004). *Gut Reactions: A Perceptual Theory of Emotion*, New York, Oxford University Press.
- Scarantino, A. and Nielsen, M. (2015). Voodoo Dolls and Angry Lions: How Emotions Explain Arational Actions, *Philosophical Studies*, 172, 2975-2998.
- Pineda, D. (2019). *Sobre las emociones*, Madrid: Cátedra.
- Roberts, R. (2003). *Emotions: An Essay in Aid of Moral Psychology*, Cambridge: Cambridge University Press.
- Scarantino, A. (2010). Insights and Blindspots of the Cognitivist Theory of Emotions, *British Journal for the Philosophy of Science*, 61(4): 729–768.
- Smith, J. (2014). Are Emotions Embodied Evaluative Attitudes?, *Disputatio*, 6(38): 93-106.
- Smith, M. (1998). “The Possibility of Philosophy of Action”, en J. Bransen (ed.) *Human Action, Deliberation and Causation*, Dodrecht: Kluwer, p. 17-41.
- Solomon, R. (1993). *The Passions: Emotions and the Meaning of Life*, Hackett Publishing Company.
- Solomon, R. (2004). “Emotions, Thoughts, and Feelings: Emotions as Engagements with the World” en R. Solomon (ed.) *Thinking about Feeling: Contemporary Philosophers on Emotions*, Oxford/New York, Series in Affective Science, p. 76-88.
- Tappolet, C. (2016). *Emotions, Values and Agency*, Oxford: Oxford University Press.
- Tietjen, R. R. (2020). The Emotion of Self-reflexive Anxiety, *Journal of the American Philosophical Association*, 297-315.
- Teroni, F. (2016). Emotions, Me, Myself and I, *International Journal of Philosophical Studies*, 24: 433-451.
- Vendrell Ferran, I. (2018). “Emoción”, *Enciclopedia de la Sociedad Española de Filosofía Analítica* (URL: <http://www.sefaweb.es/emocion/>)

Wiggins, D. (1987). *Needs, Values, Truth: Essays in the Philosophy of Value*, Oxford: Blackwell.

Wollheim, R. (1999). *On the Emotions*, New Haven and London, Yale University Press.

Zinck, A. (2008). Self-referential Emotions, *Consciousness and Cognition*, 17: 496-505.

Summary in English

1. Introduction

I have always been intrigued by the intimate aspect of our emotions. Although emotions are experiences that connect us to others in deep and meaningful ways, there is something about them that makes them, at the same time, the most intimate of our experiences. The way in which I understand the latter departs from the notions of privacy or interiority which are central to the debate on the asymmetry between the first and the third person. The notion of intimacy to which I allude here is linked, rather, to the experience of being vulnerable. As is well known, we can only feel an emotion for something we care about and, in this sense, every emotional experience is an experience in which we are exposed, that is, in which there is something at stake for us. From this perspective, inquiring about the intimate aspect of our emotions involves exploring the way in which we are tied to the things that matter to us.

Analysing the way in which our self is vulnerable to what surrounds it is a crucial task for anyone who thinks that knowing oneself is a basic requirement for a meaningful life. At least, this is what I thought when I decided to pursue a PhD in philosophy of emotions. Although emotions are closely related to a large number of elements – the body, the world, others, values, etc. – the problem of the self that expresses and exposes itself in emotions seemed to have a certain priority over other related questions. From a somewhat naive perspective, I wondered how it was possible to study the connections between our emotions and others or the world without first examining the way in which the self is involved in them. It seemed to me that the most basic feature of our emotions is that, in them, the centrality of the self is undeniable: I respond emotionally to an object because such an object has something to do *with me*, that is, it *concerns me*. And this is so regardless of what the object of the emotion is, the proportionality of the response, the truth or falsity of the beliefs involved, and so on. This conviction led me to dive into the problem of the self in emotions as a prior and, in some sense, separate question from the set of elements mentioned above.

Throughout the work I have developed in this thesis I have nevertheless come to question the idea that we can approach the self as a reality that is separate from the world, the body and others. Paradoxically, my interest in understanding the intimate aspect of emotions – where the vulnerability of the self would most clearly manifest itself – has led me to the belief that what is specific about the self that is expressed in emotions is its intimate relation or deep continuity with the world, the body and others. In a certain sense,

in my attempts to focus on the self, I have had to retrace the path that I thought would lead me to it in order to return to a series of elements – that are usually understood in opposition to the self – without which the self can be no more than a mere theoretical simplification.

In this summary, I will begin by explaining my first approach to the philosophy of emotions as well as to the question of how the self is involved in emotions. As I will briefly mention, this first approach resulted in the conviction that we need to revise the way in which the subject is usually understood in the current philosophical literature on this topic. After this, I will introduce the three theses that I defend in this work and try to show how, taken together, they aim to put into question what I call “the divided conception of the subject”. Lastly, I will summarise the content and the most important conclusions of the three papers that make up this thesis.

2. First approach to the self in emotions

In my first approach to the philosophy of emotions, I was fascinated by the work of Peter Goldie and, in particular, by his idea that emotions, qua mental states, have an indexical content that locates us, engages us emotionally with the world and prepares us for action in a special way. Following John Perry’s (1979) example in which a professor believes that he has a meeting in the afternoon but will only get up to attend it when he also believes that it is *now* afternoon – although his belief has exactly the same content before and after he gets up: “the meeting starts in the afternoon” –, Goldie introduces the following example:

Analogously, imagine you are in a zoo, looking at a gorilla grimly loping from left to right in its cage. You are thinking of the gorilla as dangerous, but you do not feel fear, as it seems to be safely behind bars. Then you see that the door to the cage has been left wide open. Just for a moment, though, you fail to put the two thoughts—the gorilla is dangerous, the cage is open—together. Then, suddenly, you do put them together: now your way of thinking of the gorilla as dangerous is new; now it is dangerous in an emotionally relevant way *for you*. The earlier thought, naturally expressed as ‘That gorilla is dangerous’, differs in content from the new thought, although this new thought, thought with emotional feeling, might also be naturally expressed in the same words. Now, in feeling fear towards the gorilla you are emotionally engaged with the world, and, typically, you are poised for action in a new way—poised for action out of the emotion (2000a, 61, emphasis added).

The first few times I read this example – about which Goldie says nothing more – I had the intuition that in it lay the key to understanding the intimate aspect of emotions. I had the impression that Goldie was trying to capture the way in which emotions, unlike other mental states, add something like “this has to do *with me*” to the thoughts that reflect our interaction with the world and others. This example seemed to pick up the intuition that emotions, in addition to being directed at the world, always make reference to oneself and one’s concerns: in feeling fear, the perception that *I am* in danger completely permeates the perception of the dangerous object in front of me and seems to explain the fact that, only when I feel the emotion, I start moving – by running away, for instance. Similarly, it is only possible to feel anger or indignation if the offensive remark I have just heard concerns me in some way, to feel shame about my actions or those of another is only intelligible if such actions have something to do with me, etc. The fact that my anger, indignation or shame should motivate me to act in a certain way would be explained precisely by my being involved in a relevant way in what is going on⁶.

Based on this reading of Goldie’s passage, I became interested in how emotions seem to locate us, that is, in how they seem to make reference to the position we are in as well as to the way in which we are involved in a situation. From this point of view, emotions have great epistemic value not only in relation to the things that surround us – fear informs me that I am facing something dangerous; humiliation, that I am facing something humiliating, etc. – but in relation to ourselves: where I am, in what ways I am affected by the object of the emotion, how I can respond to it, what my tendencies and my character are, and so on.

As I delved into the literature on the relation between emotions and the self, I began to notice that the different debates I was analysing as well as my own starting point were theoretically shaped by a certain conception of the self or the subject. Specifically, by *a divided conception of the subject* the main feature of which is that it approaches the subject through a series of oppositions: subject-world, subject-body and subject-others. What I had initially taken to be a neutral starting point appeared to correspond to a set of philosophical presuppositions about the self that, when examined closely, seem problematic and condition our reflection on the intimate aspect of emotions in significant

⁶ I can, of course, be wrong about whether a comment is offensive, whether its content is related to me, etc. The important point is that, if I do not think that the comment, the actions or whatever the object of the emotion is has something to do with me – more or less directly –, it is not intelligible for me to feel an emotion about it.

ways. This concern transformed my initial project into an attempt to explore this divided conception of the subject in order to show its problems and suggest the importance of an alternative conception that would allow us to approach the self that is expressed in emotions in a more plausible way.

3. A critique of the divided conception of the subject

In order to examine the way in which the current philosophical debate on emotions conceives of the subject and highlight the problems arising from said conception, I have examined three important contemporary debates: the debate on reflexive emotions (Cabrera *in press*), the debate on expressive action (Cabrera 2022b) and the debate on the relation between emotions and evaluative properties (Cabrera 2022a). Rather than offering a complete picture of the standard way in which the self in emotions tends to be understood, I have defended in each of these papers a thesis that reveals the problems generated by the divided conception of the self.

Thesis 1: Shame and pride are not necessarily directed at the self as the intentional object of the emotion.

Thesis 2: We are agents, not patients, of our facial and bodily expressions of emotion.

Thesis 3: Some of the problems that certain contemporary theories of emotions face seem to be rooted in a particular conception of evaluative properties as independent of the subject's emotional responses, rather than in the apparent complexity of emotions.⁷

My defence of these three theses reveals the general tendency to identify the subject with her mental life, understood as the set of cognitive processes that take place inside her head, and to treat her as an isolated reality that can be conceived of independently of the world, her body and others. In other words, the defence of these theses aims at bringing to light the way in which the following oppositions

- subject-world

⁷ As I explain below, the main goal of Cabrera (2022a) is to present a state of the art, in Spanish, on the philosophical debate on emotions. I only briefly review the dominant conception of evaluative properties and argue in favour of Thesis 3 in the last section of this paper.

- subject-body
- subject-others

systematically run through the current philosophical literature on emotions. This way of approaching the subject constitutes a divided conception because it detaches the subject from a series of realities that are fundamental to the understanding of her emotional life. As I briefly explain below, each of the debates I examine in this thesis assumes one of the different versions of the divided conception of the subject.

Firstly, the debate on reflexive emotions presupposes an artificial view according to which *subject and object can be neatly dissociated* (**the subject-world divide**). According to the arguments I review in Cabrera (in press), an emotion must be intentionally directed either at the subject or at an object different from the subject, depending on which of these two elements exemplifies the evaluative property associated with the emotion. This idea has led several authors to conclude that pride and shame are necessarily directed at oneself qua responses to how praiseworthy or shameful we are. Such an idea implies that pride and shame are not responses to what we do, *but to our own image*. Emotions such as fear, anger or joy, on the other hand, would typically make reference to objects different from the self for, in their case, the object that exemplifies the relevant evaluative property – dangerousness, offensiveness, or being worthy of celebration – is not the subject. My review of some of the arguments in favour of this view aims at showing that this way of understanding the reflexivity of shame and pride faces serious difficulties. Rather than directly exploring their commitment to the divided conception of the subject, I argue that such arguments fail to establish that shame and pride are necessarily directed at the self as the intentional object of emotion (Thesis 1) and that drawing a clear line between the features that belong exclusively to the subject and those that belong to the object of her emotion seems highly problematic.

Secondly, participants in the debate on expressive action argue that *the subject is disconnected from the gestures* through which she expresses emotions (**the subject-body divide**) and ignore the role that others – *intersubjectivity* – play in recognising such gestures as the responses of a subject (**the subject-others divide**). According to the most widespread view, we cannot approach the subject as the agent of her smiles, tears, laughter or frowns because these are non-instrumental involuntary movements that we can only explain causally. This means that facial and bodily expressions of emotion are mere happenings that the subject undergoes: they are detached from her response to the

world and her evaluative perspective. In Cabrera (2022b), I try to show that, on the basis of the main assumptions in the debate on expressive action, we are agents, and not patients, of our facial and bodily expressions of emotion (Thesis 2). My review of the main assumptions in this debate reveals the general tendency to view the body as disconnected from the mental life of the subject, which seems to explain the different treatment that the most basic emotional expressions have received compared to more sophisticated actions.

The exclusion of smiles and frowns from the category of actions, moreover, has to do with the fact that, standardly, the criterion for determining whether something is an action or a happening is whether the subject exemplifies a certain combination of beliefs and desires. Shifting our focus to the practices (that is, to intersubjectivity) in which facial and bodily expressions of emotion are embedded helps us see how problematic it is to reduce such expressions to mere happenings and to appreciate the constitutive relation or the deep continuity that exists between the subject, her body and others.

Thirdly, the current debate on emotions tends to conceive of *evaluative properties as independent of the subject and her emotional responses* (**the subject-world divide**). It is generally assumed that the different types of emotion are individuated on the basis of different evaluative properties – fear is a response to what is dangerous; anger, to what is offensive; humiliation, to what is humiliating, etc. – and that evaluative properties supervene on natural properties and are individuated independently of the emotional responses of the subject. This idea has led certain contemporary theories of emotions to propose different explanations of the way in which emotions register the evaluative properties that are exemplified by the object of the emotion. As Deonna and Teroni put it, “intuitively, no other emotional experience than that of fear is a suitable candidate for presenting the world in terms of a danger” (2012: 86). Theories that assume this view of evaluative properties, however, tend to stall in their attempts to explain how emotions can give us access to a kind of properties that are entirely independent of them. In Cabrera (2022a) I briefly review this philosophical commitment as well as some of the arguments that have been presented against the contrary view, that is, the view that holds that emotional responses and evaluative properties are bound by a relation of mutual dependence. Such a review does not aim to establish which of the two conceptions of the relation between emotions and evaluative properties is preferable, but rather to show that it is not obvious that we should reject the conception of the mutual dependence in favour of the dominant view. This review suggests, moreover, that some of the problems that

certain contemporary theories of emotions face may be rooted, not so much in the apparent complexity of emotions, but in their commitment to the conception of evaluative properties as independent of the subject's emotional responses (Thesis 3).

As we can see, these three debates are shaped by a divided conception of the subject, which constitutes a problematic and impoverished conception for several reasons. Firstly, when we oppose the subject to the objects of her shame and pride, the subject ends up losing part of the intimate bond that connects her to such objects and becomes disproportionately self-centred. This movement, moreover, disconnects her from others in a relevant sense. Secondly, by detaching the subject from her body and from others, the subject becomes a poor victim of a series of facial and bodily changes that have nothing to do with her. And, thirdly, when the subject is cut off from the evaluative properties to which she responds, the subject is reduced to a mere detector of properties and her essential role in the creation of meaning and values disappears. The importance of reviewing this conception of the subject lies, moreover, in the fact that it is a conception that has a great influence on the current study of emotions since it determines the kind of questions that guide our research. As we will see in more detail below, within the debate on reflexive emotions, the subject-world divide has led to a characterisation of shame and pride that assigns the self a very implausible place. Within the debate on expressive action, on the other hand, the subject-body and subject-other divides have resulted in an over-intellectualised view of our agency that is unable to explain all of the things that we do as agents. And finally, within the debate on the relation between emotions and evaluative properties, the subject-world divide has led several recent theories of emotions to stall in their efforts to explain the way in which emotions give us access to evaluative properties.

For these reasons, my defence of the above three theses suggests, moreover, that a *more integrated conception of the subject* that takes into account its *constitutive relation* with the world, her body and others would be preferable to the divided conception just described. As is evident, a more integrated conception would set a different direction both in the debates analysed here and in the philosophy of emotions in general. This issue, however, exceeds the limits of this thesis and will be a project for another time. In what follows, I set out in detail the content and the most important conclusions of each of the papers that make up this thesis.

4. On the Intentionality of Shame and Pride

In the first paper (Cabrera in press), I review one of the most widespread assumptions in the current debate on reflexive emotions. According to a significant number of authors, the main difference between shame and pride, and the rest of our emotions – fear, hatred, surprise, joy, jealousy, anger or compassion – is that the former are necessarily directed at the self as the intentional object of the emotion, whereas the latter are typically oriented towards objects in the world other than the self (Helm 2001, Zinck 2008, Deonna and Teroni 2012, Deonna, Rodogno and Teroni 2012, Teroni 2016 and Tietjen 2020). In this paper I attempt to show that the arguments provided in favour of this idea fail to show that shame and pride are necessarily directed at oneself – *the doer* – rather than at an object different from the self – *the deed*. I suggest, moreover, that a satisfactory account of these emotions should be able to account for the normatively paradigmatic cases in which the subject of shame and pride focuses her attention on what she has done rather than on herself.

In order to show that shame and pride are not necessarily directed at the self as the intentional object of emotion (Thesis 1), I review three arguments in favour of the most widespread view and offer a series of counterarguments. As we shall see, proponents of this view seem, on the one hand, to assign the self the wrong place in shame and pride and, on the other, to rely on a problematic assumption according to which the features of the object of an emotion can be neatly differentiated from the features of the subject who experiences it.

4.1. The Correctness Argument

Firstly, according to The Correctness Argument (CA), shame and pride are necessarily directed at the subject of the emotion because, in order to be assessed as correct emotions, the subject must exemplify the evaluative property that constitutes the formal object of the emotion, unlike what happens in the case of non-reflexive emotions (Teroni 2016, 5). For instance, the shame I feel after telling a lie to a friend of mine is correct if *I* am an unworthy or shameful person. On the other hand, the fear I feel towards the dog in front of me is correct if *the dog* is dangerous. As we can see, in the first case, the evaluative property – unworthiness or shameful – is exemplified by the subject of the emotion whereas, in the second, the evaluative property – dangerousness – is exemplified by an object different from the subject. It follows from this, according to CA, that I am the

intentional object of my shame and that the dog is the intentional object of my fear. This argument thus seems to assume that the object that provides the criterion of correctness of an emotion – that is, the object that exemplifies the evaluative property in question – *is the object* of the emotion.

This argument, however, ignores the fact that, in the case of shame and pride, there are always two objects that can exemplify the relevant evaluative property. When I am ashamed of the lie I have told my friend, not only I but *what I have done* – lying and harming my friend – constitutes something unworthy or shameful. Similarly, when I am proud of my business, not only I but the business itself is something praiseworthy. What I do, however, seems to have an important priority in determining the correctness or proportionality of my emotional response: the fact that telling a lie to my friend and harming her is something unworthy is, on the one hand, what makes me an unworthy person (and not the other way around) and, on the other, what determines that I should be ashamed of what I have done. This means that, although in the case of shame or pride the subject exemplifies the evaluative property in question, this *does not imply that she is responding emotionally to herself*.

This line of reasoning, rather than definitively determining what the object of shame and pride is – myself or what I have done – shows how difficult it is to distinguish neatly between the features of the subject and the features of the object of her emotion. The evaluative properties that determine the correctness of an emotional response seem to depend on both the subject and the object to which she responds emotionally, so any conception that insists that the evaluative properties have to be exemplified either by the subject or by the object of the emotion will not be able to offer us a very plausible analysis of these emotions.

4.2. The Phenomenological Argument

Secondly, the Phenomenological Argument (PA) claims that the fact that shame and pride are necessarily directed at the self is manifested in the way we focus our attention on ourselves when we feel such emotions, as opposed to what happens when we feel fear, joy, compassion, and so on. (Teroni 2016, 6). That is, given that when I feel shame my thoughts and attention revolve around myself – “what is everyone going to think of me now?”, “how could I have done this?”, “I should have known better!” –, it seems evident that my shame is about myself, that is, that it is intentionally oriented towards myself.

This argument, however, ignores cases of shame and pride in which the subject focuses on what she has done rather than narcissistically or fearfully turning towards herself and her image (Moran 2001, Corbí 2012). There seems to be a relevant difference between focusing my attention on myself when feeling ashamed about the lie I have told my friend – on what kind of person I am, what everyone will think of me, etc. – and focusing my attention on my friend and the harm I have caused her – on what she needs from me now, how to repair the harm, etc. Generally, we censor the first kind of cases as inappropriate or disproportional responses to the situation in which we find ourselves: although we tend to care about our image, this does not seem to be the way we *should* respond when we have harmed someone. This means that the examples of shame in which the subject focuses on the harm she has caused are *normatively paradigmatic cases* that allow us to criticise the cases in which the subject focuses on herself as deviations. The fact that there is a psychological tendency to respond in a deviant way when we feel shame or pride, however, *does not imply that the subject is the intentional object of the emotions in question.*

4.3. The Grammatical Argument

Lastly, according to The Grammatical Argument (GA), shame and pride are necessarily directed at oneself because we can always insert an explicit reflexive component in sentences that refer to these emotions, unlike what happens with non-reflexive emotions (Teroni 2016, 5). The idea here is that we can always transform “Mary is proud of the canvas” into “Mary is proud *of herself* because she painted this canvas” and “Mark is ashamed of his big nose” into “Mark is ashamed *of himself* because of his big nose” without relevant changes in the meaning of these sentences.

Regarding the first sentence, however, it seems clear that its meaning changes significantly after the transformation: at first, Mary’s pride seems to be intentionally directed at the canvas she has painted while, after the transformation, her emotion seems to be directed at some skill or trait of hers: being a good painter, having benefited from the painting classes she has received, etc. We may wonder, furthermore, whether feeling an emotion towards a trait of ours – our painting skills, the size of our nose, etc. – is equivalent to feeling an emotion towards ourselves, as GA assumes. It seems, however, that being ashamed of myself is something much deeper than being ashamed of any of my features: while the shame I may feel about a certain part of my body may more or less

superficially affect my behaviour and concerns, valuing myself globally as a shameful or unworthy person clearly has a much more significant impact on my life.

This means that we do not seem to be able to carry out the transformations proposed by GA without important changes in the intentional object of shame and pride. Moreover, and just as our review of CA showed, the analysis of this argument seems to reveal how problematic it is to draw a sharp line between the features that belong exclusively to the subject and the features that are alien to her.

4.4. Conclusions

Through this brief review I have tried to show that the proposed arguments do not succeed in showing that shame and pride are always intentionally directed at the self, while the rest of our emotions are oriented towards objects in the world other than the self. It seems, therefore, that anyone who wants to insist on this conception of the reflexivity of shame and pride should provide new arguments in its favour. It follows from this discussion, moreover, that a satisfactory account of the reflexivity of such emotions should be able to explain the cases that we have called normatively paradigmatic. Otherwise, we will continue to approach the self in the emotions exclusively by reference to a series of cases that do not allow us to study shame and pride in all their complexity.

As this analysis reveals, the contemporary debate on the reflexivity of shame and pride seems to be shaped by the subject-world divide, which presupposes that it is possible to draw a neat line between the features of the subject and the features of the objects to which she responds. This divide seems to generate important problems since, on the one hand, it has given rise to a very implausible view of the intentionality of shame and pride: as the normatively paradigmatic cases we have discussed show, the subject does not always seem to be responding emotionally to herself when she feels such emotions (and, when she focuses on herself, we censure her responses as inappropriate). On the other hand, it is a philosophical commitment that has led us to focus on such artificial questions as whether the relevant evaluative property is exemplified by the subject (the doer) or by the object of her emotion (the deed). Perhaps, the evaluative property of being unworthy or shameful applies, rather, to a complex whole of which both the doer and the deed are part, and there is no criterion that could allow us to distinguish between the different elements in order to determine which of them really or primarily exemplifies the evaluative property in question. In any case, and beyond these last speculations, it seems reasonable to conclude that this review of the debate on reflexive emotions points to the

need to find an alternative view to the divided conception of the subject presupposed in such discussion.

5. Duchenne Smiles are Actions not Mere Happenings: Lessons from the Debate on Expressive Action

In the second paper (Cabrera 2022b), I argue that, contrary to what is often assumed in the debate on expressive action, we are agents, and not patients, of our facial and bodily expressions of emotion (Thesis 2). To this purpose, I compare facial and bodily expressions of emotion – smiling, frowning, crying, etc. – with simple expressive actions – jumping for joy, covering one's face in shame, scratching one's head in frustration, etc. – and I argue that, on the basis of the main assumptions in the debate, both kinds of behaviour should be classified in the same way, that is, either as actions or as mere happenings. The latter, however, conflicts with the way we relate to behaviours such as jumping for joy or covering one's face in shame, therefore, I conclude that both facial and bodily expressions of emotion and simple expressive actions should be included within the category of actions.

5.1. Facial and bodily expressions of emotion

Within the debate on expressive action, there is a wide consensus with respect to the idea that facial and bodily expressions of emotion – smiling, frowning, crying, clenching one's fists, hanging one's head, etc. – are not actions of an agent but mere happenings that she undergoes (Hursthouse 1991, Goldie 2000b, Betzler 2007 and 2009, Helm 2001 and 2016, Bennett 2016, Müller and Wong forthcoming). The criteria that have been provided in order to justify the exclusion of these kinds of behaviours from the category of actions are the following: facial and bodily expressions of emotion (C1) are not instrumental behaviours, (C2) are alien to rationality, and (C3) are involuntary. According to this, then, smiles and frowns are analogous to spasms and reflexes qua mere happenings that are disconnected from our agency. From this perspective, smiles are not something *we do* but rather something that *happens to us*.

This widespread view, however, generates an important dilemma within the debate on expressive action, for one cannot claim that facial and bodily expressions of emotion are not actions while arguing that simple expressive actions are, since, as will be argued below:

- Neither can be instrumentally explained (they both satisfy C1)
- Both can be rationally explained (neither of them satisfies C2)
- Both can involve muscles that we cannot directly control (they both can satisfy C3)

The fact that both kinds of behaviour satisfy exactly the same criteria implies that *either both are actions or neither is*. The latter, however, seems counterintuitive, since we would have to assimilate simple expressive actions such as jumping for joy or covering one's face in shame to mere happenings, which conflicts with the way we relate to these kinds of behaviours. For this reason, I conclude that both facial and bodily expressions of emotion and simple expressive actions should be included within the category of actions. Before we can reach this conclusion, however, we must briefly analyse what expressive actions are.

5.2. Simple expressive actions

As Hursthouse (1991) noted, actions such as jumping for joy or covering one's face in shame pose an important problem for the standard account of agency. According to this view, a behaviour can only be considered an action if it can be rationalised in terms of a desire and an instrumental belief (Davidson 1963). When I run away from a predator, for instance, my behaviour constitutes an action – and not a mere happening – because we can rationalise it in terms of my desire to protect myself and my belief that the best way to protect myself is to run away. Jumping for joy or covering one's face in shame, however, cannot be rationalised in this way, for there is nothing that the agent is pursuing through her behaviour. These behaviours, nevertheless, are clearly actions, says Hursthouse – in particular, and since they cannot be instrumentally rationalised, they are *arational actions*: actions that are alien to rationality.

This proposal, however, generated a huge debate for the idea that expressive actions are disconnected from rationality did not seem convincing to a significant number of authors. When we see someone jumping for joy when receiving good news, for instance, we see that what she does *makes sense* given the emotion she is feeling, that is, her behaviour is intelligible to us. This idea has led many authors to propose different ways in which we can rationalise this kind of behaviour while preserving, at the same time, the standard account of agency (Smith 1998, Goldie 2000, Helm 2001 and 2016,

Betzler 2007 and 2009, Döring 2003, Scarantino and Nielsen 2015, Bennett 2016 and 2021, Pineda 2019, Müller and Wong forthcoming).

On the one hand, authors such as Smith (1998) or Bennett (2016) have tried to provide instrumental explanations of expressive actions in order to show that they can be rationalised in terms of the standard account. This kind of explanation, however, leads inevitably to the over-intellectualisation of simple expressive actions. Betzler (2007 and 2009) and Goldie (2000), on the other hand, have suggested that, perhaps, jumping for joy, scratching one's head in frustration or covering one's face in shame – since they do not exhibit the rationality of more complex actions – are not actions but mere happenings. This proposal, however, seems to have rather implausible consequences: if jumping for joy is nothing more than a mere happening, we need an explanation of the way we normally approach such behaviour. That is, we need an explanation of why, when we see someone jumping for joy, we relate to her behaviour as a response to something that the subject evaluates positively and that can be assessed as appropriate or inappropriate, unlike what happens when we see someone undergoing a spasm.

Such problems have led Helm (2001 and 2016) and Müller and Wong (forthcoming) to question the idea that for a behaviour to be considered rational, such behaviour must be rational from an instrumental point of view.

- *Non-instrumental Rationality Assumption:* jumping for joy – as well as the rest of our simple expressive actions – can be considered an action because it has *a point* that shows that the agent is responding to reasons through her behaviour and that, therefore, such behaviour is not accidental.

On the basis of this assumption, then, it seems that we can preserve the intuition that simple expressive actions *are actions* and avoid their over-intellectualisation.

5.3. Facial and bodily expressions of emotion vs. simple expressive actions

Once we have made room in the debate for simple expressive actions, we are in a position to argue – through arguments A1-A3 – that, on the basis of C1-C3, both simple expressive actions and facial and bodily expressions of emotion should be included within the category of actions.

(A1) If we accept that simple expressive actions are actions despite their not being instrumental, we cannot exclude facial and bodily expressions of emotion from the category of actions on this ground (as C1 claims).

As we have just seen, if we accept the *Non-instrumental Rationality Assumption*, simple expressive actions can be considered actions even though the agent does not aim at satisfying some end we she performs them. This means that we cannot exclude facial and bodily expressions of emotion from the category of actions just because they lack an instrumental structure.

(A2) If an action can be rational just because it has a point without being instrumentally rational, we have no reason to exclude facial and bodily expressions of emotions from the category of actions (contra C2).

Although participants in the debate on expressive action unanimously consider that facial and bodily expressions of emotion are alien to rationality (Helm 2001 and 2016, Bennett 2016 and Müller and Wong forthcoming), a close look at the practices in which such expressions are embedded reveals that this cannot be correct, for (a) we relate to facial and bodily expressions of emotions as responses to something and (b) we regard them as revealing of the subject's evaluative perspective on her situation. Moreover, (c) they seem to be regulated by moral considerations and (d) they can be educated. If facial and bodily expressions of emotion were really mere happenings, the practices just described would be unintelligible.

(A3) If actions must be voluntary, then facial and bodily expressions of emotion are not actions, but neither are they many instrumental behaviours that, according to the standard account of agency, should be considered actions (contra C3).

As Goldie points out, when our smiles are genuine, we cannot directly control all the muscles involved in them (2000b, 35). This shows, he argues, that smiles are not actions but mere happenings that we undergo. However, there are many instrumental behaviours – which, according to the standard account of agency, are clearly actions – in which we cannot directly control all the muscles of our body either: when I run away from a predator, due to the adrenaline in my system, the way I move my body is different from the way I can move my body when, for instance, going for a jog. This fact, nevertheless, does not prevent us from considering that what I do – running away from the predator – is an action of mine and not a mere happening. Therefore, it does not seem correct to exclude facial and bodily expressions of emotion from the category of actions just because we cannot directly control all the muscles involved in such behaviours.

5.4. Conclusions

Through arguments A1-A3, I have tried to show that, on the basis of criteria C1-C3, facial and bodily expressions of emotion cannot be classified differently from simple expressive actions. Since assimilating the latter to mere happenings is highly problematic, I have concluded that both kinds of behaviour should be considered actions. It follows from this that we are agents, and not patients, of our facial and bodily expressions of emotion.

The general tendency to treat facial and bodily expressions of emotion differently from the rest of our actions reveals the strong influence that the subject-body divide has within the debate on expressive action. The idea that smiling, crying or frowning are not expressions of our agency presupposes a view of the body as something separate from the subject – as something that, at most, reveals the existence of a subject through a series of symptoms but that cannot be identified with it. As we have seen, this way of conceiving facial and bodily expressions of emotion is highly implausible even on the basis of the very premises of the debate on expressive action.

Furthermore, both participants in this debate and the standard account of agency approach the subject as a reality that is detached from others and examine the question of her agency by appealing exclusively to the exemplification of a series of mental states⁸. When we look at the subject in her connection to others, however, it becomes clear that her smiling, laughing and crying cannot be mere happenings that she undergoes: the practices that regulate such expressions reveal that they are the responses of a subject and that, therefore, the commonly held view in the debate on expressive actions is clearly misguided. The most important conclusion that we can draw from this review is that it seems that a good account of facial and bodily expressions of emotion should approach the subject from the constitutive relation that she has with her body and others.

6. *¿Qué es una emoción?*

The aim of the third paper (Cabrera 2022a) is to present a state of the art on the philosophical debate on emotions⁹ and to carry out a brief review of the philosophical commitments that may be influencing some of the most important theories of emotions.

⁸ With the notable exception of Helm (2001 and 2006), Bennett (2021) Müller and Wong (forthcoming).

⁹ Unlike in the English-speaking context, where encyclopaedia entries and handbooks about the philosophy of emotions abound, there are few materials in Spanish that provide an overview of the state of the philosophical research on emotions. The recent works by Vendrell (2018) and Pineda (2019) are an example of the growing interest in making this field accessible to readers in Spanish.

In particular, I examine the widespread commitment to a conception of evaluative properties as independent of the subject's emotional responses and argue that the problems that certain contemporary theories face may be rooted in such a conception rather than in the apparent complexity of emotions (Thesis 3). We begin with some brief points of clarification and then examine the different theories.

6.1. Initial considerations and requirements for a good theory

Firstly, this paper approaches the study of emotions through a personal kind of explanation. In order to account for our normative talk about emotions – “you shouldn't feel guilty about this”, “the comment he made was very offensive, of course she is angry”, etc. –, our explanations must accommodate the fact that the subject has a certain *point of view* on the situation in which she finds herself and that it is her, and not merely her body or a series of causal processes that are alien to her evaluative perspective, who responds emotionally. Although all of the theories we examine in this paper seek to account for this kind of talk, we will see how some of them do not seem to meet this aim. Secondly, we reflect on emotions understood as *episodes* and not as dispositions. Although it is not clear that emotions can be reduced to mere emotional episodes (see Wollheim 1999 and Goldie 2000a), it is commonly assumed that emotions are episodic states of short duration that are felt in a certain way.

Within the current debate, it is also commonly presumed that a satisfactory theory must be able to explain three fundamental aspects of emotions. In the first place, emotions are intentional phenomena: just as memories, perceptions or beliefs, emotions are directed at or are responses to objects, situations, events or states of affairs. Fear, for instance, is always fear of *something* and not a mere brute feeling. Secondly, emotions are evaluative phenomena, that is, they always represent the object to which they are directed under a certain evaluative perspective (Kenny 1963). This means that fear is a response to what is dangerous; humiliation, to what is humiliating; joy, to what is worthy of celebration, and so on. This allows us to identify the different types of emotion qua responses to different evaluative properties. Finally, emotions have a felt quality, which means that they are *experiences* with a certain phenomenology. Some have tried to explain this felt quality exclusively in terms of perceptions of bodily changes (James 1884, Deonna and

Teroni 2012) while others have tried to describe more “experiential/lived”¹⁰ feelings that are not only about how the body feels (Goldie 2000a, Deigh 2010, Vendrell 2018).

6.2. The belief-desire theory

The belief-desire theory identifies emotions with combinations of beliefs and desires: the fear we may feel for a dog would consist in the belief that the dog is going to attack us and the desire to run away, and the regret we may experience for having lied to a friend would consist in the belief that we have been dishonest and the wish that we had behaved differently (Gordon 1987, Green 1992. See also Davidson 1976). This theory also holds that the bodily changes associated with emotions are mere causal effects of these combinations of beliefs and desires and reduces the felt aspect of emotions to the perception of such changes.

This way of understanding emotions, however, has been widely rejected on several grounds: (a) bodily phenomenology does not seem to exhaust the way emotions are felt, (b) combinations of beliefs and desires are neither necessary nor sufficient to have emotions, (c) nor do they allow us to distinguish between different types of emotion and (d) the possibility of feeling emotions is reduced to language possession, since the ability to have beliefs and desires presupposes a propositional kind of thought. According to this last objection, the main assumptions of the belief-desire theory imply that animals and young children cannot feel emotions, which is clearly incorrect (Deigh 2010). Although this last idea is often presented as a devastating objection to the belief-desire theory, the fundamental assumption on which it is based – that the ability to have beliefs and desires depends on language possession – is quite philosophically controversial (Dennett 1995).

6.3. The judgmentalist theory

The judgmentalist theory, on the other hand, holds that emotions are a particular kind of evaluative judgement about objects and situations that concern us: fear of a dog would consist in judging that the dog is dangerous, and regret for having lied to a friend would consist in judging that the lie is something regretful (Solomon 1993, Nussbaum 2001). By including evaluative concepts in the judgements that constitute emotions, this theory allows us to individuate the different types of emotion (Deigh 1994, Roberts 2003).

¹⁰ *Vivencial* in the original Spanish.

However, and in line with the previous theory, judgmentalism holds that the bodily changes associated with emotions are mere causal effects – and not a constitutive part – of emotions, which means that the felt aspect of emotions is also reduced here to the perception of these changes. This is one of the reasons why this theory has received numerous objections.

Critics of this theory have also pointed out (a) that evaluative judgements do not seem to be necessary nor sufficient to experience emotions, (b) that the explanation of recalcitrant emotions provided by judgmentalism is unsatisfactory, and (c) that the capacity to have emotions is reduced to the possession of evaluative concepts and, therefore, of language. This last objection, however, as we have seen, is based on a fairly controversial assumption. In their attempts to qualify the kind of judgement that emotions are (Nussbaum 2001, Solomon 2004), advocates of the judgmentalist theory have also been accused (d) of introducing an *ad hoc* kind of judgement in order to avoid the objections raised against their theory (Scarantino 2010). This kind of accusation, however, seems to presuppose a very specific notion of judgement as paradigm. Although judgmentalism is often accused of over-intellectualising emotions, critics of this theory themselves seem to hold an over-intellectualised view of judgements. Perhaps the idea that emotions are a kind of judgement would allow us to modify our concept of judgement if we discover that it is a concept that can only account for a number of highly intellectualised cases.

6.4. The perceptual theory

The perceptual theory, by contrast, identifies emotions with perceptions of values or evaluative properties (De Sousa 1987, Roberts 2003, Tappolet 2016). According to this theory, just as we perceive the blueness or the roundness of a ball, we perceive the dangerousness of a dog or how regrettable it is to lie to a friend. Proponents of the perceptual theory argue that perception provides a good model for understanding emotions because both kinds of states are intentional, do not require the use of concepts and have a felt quality that remains unexplained if they are reduced to mere judgements. Although this theory seemed to be a promising alternative to judgmentalism, it is based on a very controversial view of how perception works. In particular, it is not clear how perception understood as a passive, non-conceptual registration of our surroundings could ever justify the beliefs and judgements that we form based on our perceptual experiences (McDowell 1994).

Both advocates of the perceptual theory and its objectors, however, presuppose this model of perception, which means that the objections raised against this theory will be convincing depending on whether or not we are sympathetic to such model. Such objections strive to show that there are deep differences between emotions and perceptions: (a) emotions require intentional bases in order to access their object while perceptions access their object directly, (b) emotions have a “response” phenomenology and perceptions a “presentational” one, and (c) emotions, unlike perceptions, cannot be directed at properties but are directed exclusively at the bearers of such properties (Deonna and Teroni 2012, Mitchell 2021).

6.5. The Jamesian theory

According to the Jamesian theory, emotions are perceptions of bodily changes: fear of a dog would consist in feeling our hearts race, our bodies freeze, our hands shake, etc., while regret for having told a lie would consist, for instance, in noticing our head drop, a knot in our stomach, etc. (James 1884). Although, unlike the belief-desire theory and the judgmentalist theory, the Jamesian theory assigns a major role to the felt aspect of emotions, such theory faces two fundamental objections: (a) the first is that there seem to be people who, due to spinal cord injury, cannot feel their bodies and yet have emotions (Cannon 1927, Cobos et al. 2002) and (b) the second is that this theory attributes the wrong kind of intentionality to emotions (Pineda 2019). If emotions are perceptions of bodily changes, then such changes constitute their intentional object. However, as the normative talk that regulates our emotions shows, emotions are not responses to changes that take place in our bodies but to objects, people, events, etc.

Prinz’s (2004) neo-Jamesian theory attempts to avoid this second objection by adopting a teleosemantic conception of intentionality according to which the bodily changes associated with emotions represent evaluative properties. The idea here is that the bodily changes that are brought about by fear – trembling, sweating, etc. – would have been evolutionarily selected as indicators of the presence of danger and, therefore, have the function of carrying this information. However, there seems to be no place for a subject who evaluatively apprehends, from a first-person perspective, the object of her emotion in this form of representation (Deigh 2010). Moreover, the primitive intelligibility that seems to exist between the phenomenology of emotions and the corresponding evaluative properties is lost: intuitively, the way the body feels does not

seem to be a mere symptom of the presence of an evaluative property (Deonna and Teroni 2012).

6.6. The attitudinal theory

The attitudinal theory, in turn, argues that emotions are bodily felt evaluative attitudes that cannot be reduced to any other kind of mental state (Deonna and Teroni 2012, 2020). Specifically, emotions are bodily stances – or action tendencies (Frijda 2007) – oriented towards the object of the emotion. Fear, for example, would consist in feeling the body preparing itself to lessen the impact of the dog that is about to attack us. Shame, on the other hand, would consist in feeling the body prepare to disappear into the ground or from the view of others¹¹. According to this theory, we evaluate the object of our emotions through the way our body is poised to act in relation to it. The attitudinal theory, however, (a) is vulnerable to the objection about the people with a spinal cord injury who, despite not feeling their bodies, still have emotions (Cannon 1927, Cobos et al. 2002), (b) fails to explain the correspondence between behavioural patterns and evaluative properties, (c) claims that it is the body – and not the subject – that is poised to act, (d) and ignores the fact that our normative talk about emotions does not regulate or deal with bodily stances.

All these problems seem to be related to the reduction of the felt quality of emotions to how the body feels, that is, to bodily phenomenology. In the face of such difficulties, proponents of another family of theories – affective intentionality theorists – argue that we need to include other kinds of feelings or sensations that go beyond the merely bodily ones in order to understand the way in which emotions are directed at their object and give us access to their evaluative properties. For instance, through the felt approval or disapproval (Mitchell 2021) or the feelings of pleasure or displeasure (Helm 2002) associated with our emotional experiences we could gain access to how the object of our emotion is evaluatively. However, it is not at all clear how such basic affective sensations as approving or disapproving, feeling pleasure or displeasure, could give us access to the set of evaluative properties that correspond to the different types of emotion.

¹¹ Although we have so far talked about regret in our examples and not about shame, Deonna and Teroni do not explain what the bodily stance corresponding to regret might consist in. I have chosen an example from the authors in order not to distort their theory.

6.7. The problem of the relation between emotions and evaluative properties

The review of some of the most important theories of emotions shows that both the theories themselves, as well as some of the objections raised against them, assume certain philosophical commitments that seem to condition the study of emotions in significant ways. In particular, we have pointed out that the idea that language possession is necessary for the attribution of beliefs and desires is philosophically controversial, that critics of the judgmentalist theory might rely on an over-intellectualised conception of judgements and that the view of perception as a passive, non-conceptual registration of what surround us presupposed both by proponents and critics of the perceptual theory is problematic. Beyond these issues, however, there seems to be a deeper and more complex philosophical commitment that seriously conditions the proposals of the perceptual theory, the attitudinal theory and affective intentionality theories.

As is clear from our discussion, these last three theories identify the reflection on the evaluative aspect of emotions with the question about the way in which *emotions register or give us access to the various evaluative properties* that allow us to individuate them. All these theories, however, seem to get stuck on the same point. The perceptual theory, in the first place, fails to establish an analogy between the way we perceive non-evaluative properties (blue) and evaluative properties (dangerousness). Secondly, the attitudinal theory posits a very implausible correspondence between the different bodily stances associated with each emotion and the corresponding evaluative properties. And finally, with respect to affective intentionality theories, it is not clear how such basic affective sensations – approving or disapproving, feeling pleasure or pain – could give us access to complex evaluative properties such as being dangerous or reproachable.

This common problem might have to do with the apparent complexity of emotions, however, it is a problem that only arises if one is committed to *a conception of evaluative properties as independent of the subject's emotional responses*. According to the dominant view in the current debate on emotions, evaluative properties are properties of the world that our emotions register and that supervene on natural properties (Deonna and Teroni 2012, Tappolet 2016, Mitchell 2021): the property of being dangerous, for instance, would supervene on the natural property of having sharp fangs or a loud growl, and the property of being humiliating would supervene on the natural property of being ignored or called by a name that is not your own (Kim 1993).

One of the fundamental problems of this conception of evaluative properties is that we seem to need an explanation of the gap that appears to exist between the natural and the evaluative (Smith 2014): fear is not a response to what has the property of biting, but to what has the property of being dangerous, and humiliation is not a response to someone who does not call you by your name, but to someone who humiliates you. No matter how many evaluative properties we add, we will not be able to explain why a certain response is a case of fear, humiliation or any other emotion unless we make reference to evaluative concepts (Berlin 1958, Corbí 2012).

This kind of problem has given rise to an alternative conception of evaluative properties according to which emotional responses and evaluative properties are bound by a relation of *mutual dependence* (Wiggins 1987, Goldie 2000a, Dunn 2006, Corbí 2012). The main idea is that we cannot individuate evaluative properties without reference to the different types of emotion (and not just the other way around): to be offensive is to merit anger, to be dangerous is to merit fear, to be shameful is to merit shame, etc. and not something that supervenes on certain natural properties. Although this is a widely rejected conception in the contemporary debate on emotions, the arguments that have been provided against it do not seem to be definitive or, at least, do not seem to establish that a conception of evaluative properties as independent of our emotional responses is clearly preferable. Adopting one of these two positions, however, significantly determines the kind of questions that guide research on emotions, so it is pertinent to briefly review the arguments offered against the idea that there is a relation of mutual dependence between emotions and evaluative properties, as well as some of the responses offered by its advocates.

*(A1) The conception of the mutual dependence offers a circular explanation:
something has the property of being praiseworthy if it merits pride, and
something merits pride if it has the property of being praiseworthy.*

In response to this kind of objection, Wiggins points out that the kind of circularity involved in this proposal is a virtuous one (1987). When we try to elucidate what evaluative properties are, we have to look at them in the context in which they make sense. This means that emotional responses and evaluative properties have to be viewed in light of each other: in the appropriate context, the questions “should I admire x?” and “is x admirable?” are equivalent (Dunn 2006).

(A2) If evaluative properties are to be individuated by the emotional responses they elicit, I can never be wrong about whether something has a certain property.

Advocates of the conception of the mutual dependence, however, do not hold that evaluative properties are to be individuated by the *actual* emotional responses of a subject. According to Wiggins, through a process of mutual instruction and enhancement, we learn that the objects and situations to which we respond with a particular emotion paradigmatically possess a particular evaluative property. Once we have learned this, we can check whether something has a certain property by looking at the marks of paradigmatic cases and not merely at whether or not a particular subject experiences a certain emotion.

(A3) The idea that evaluative properties depend on emotional responses leads inevitably to moral relativism.

In response to this argument, it is useful to clearly differentiate moral relativism from *moral pluralism* (Berlin 1958, Corbí 2012): assuming that there is more than one valid answer to a given moral problem does not imply that any answer is valid. For a response to count as a strictly moral one, it must be linked to morally relevant aspects of the situation.

(A4) Unless the conception of the mutual dependence clarifies the kind of reasons that determine the emotion that a situation merits, such a proposal is untenable. A vile demon may not be admirable but prudentially merit admiration, and a joke may be objectively funny but not merit amusement given our moral standards.

The important question here is whether, in the first case, advocates of the mutual dependence are forced to conclude that the demon is admirable. And, as we have seen, it seems that they can acknowledge situations in which a certain subject responds with a particular emotion even if the object of her emotion does not merit such a response (Wiggins 1987). In the second case, if all that mattered was not to hurt someone else's feelings, we seem to be dealing with prudential rather than with moral considerations: even if it is correct to feel amusement, I should not laugh in front of the person who the joke is about. The reason why it is morally inappropriate to experience the joke as funny, however, seems to be rather that such an emotional experience makes me insensitive to

the harm that may be caused to the person involved and, in this respect, we might ask in what relevant sense we should continue to insist that feeling amusement is objectively correct but morally inappropriate.

6.8. Conclusions

Far from establishing which of the two ways of understanding the relation between emotions and evaluative properties is preferable, the aim of this last review has been to show that this is an open question and that it is not obvious that we should reject the conception of the mutual dependence in favour of the dominant position. As we have seen, however, several of the most recent theories of emotions are guided by the idea that evaluative properties are independent of emotions, and that part of what we need to establish when we try to understand what emotions are is how emotions register or give access to such properties. Given the difficulties that the above theories encounter in accounting for the latter, we cannot help but wonder whether this difficulty is inherent to the complexity of the phenomenon or whether it might be related, rather, to the conception of evaluative properties to which such theories are committed.

As we can see, the dominant conception of the relation between emotions and evaluative properties is shaped by the subject-world divide since, from this view, properties such as being humiliating or admirable are independent of the subject's emotional responses. Conceiving our world as something that is independent of us is, on the one hand, a major philosophical commitment and, on the other hand, a commitment that, as we have seen, is not without its difficulties. Moreover, the fact that several of the current theories of emotions stall in their attempts to explain the way in which emotions give us access to evaluative properties might indicate – as advocates of the conception of the mutual dependence suggest – that a more integrated conception of the subject with her world would allow for a better understanding of her emotions.

7. Conclusion

The theses that I have defended in these three papers reveal how the divided conception of the subject – in its different versions – conditions contemporary research on emotions in relevant ways. Regarding the debate on reflexive emotions, we have seen how the subject-world divide has led to a very implausible conception of the reflexivity of shame and pride and a quite artificial distinction between the features of the subject and the

features of the object of her emotion. In relation to the debate on expressive action, the subject-body and subject-other divides have contributed to a view of our actions that prevents us from accounting for all of the things that we do as agents. And, within the debate on the relation between emotions and evaluative properties, the subject-world divide has favoured a way of approaching the study of emotions that may be seriously misguided.

Such considerations should, at the very least, encourage us to revisit the conception of the subject to which we are implicitly committed when reflecting philosophically on emotions. These considerations further suggest the importance of providing a more integrated conception of the subject. Such a conception would have to take into account the constitutive relation or the deep continuity that seems to exist between the subject, the world, her body and others. As we said at the beginning of this summary, a more integrated conception would raise different questions and problems than those raised by the divided conception we have examined here, and would thus mark a different direction both in the debates we have analysed and the philosophy of emotions in general. This issue, however, exceeds the limits of this thesis and will be a project for another time.

References

- Bennett, C. (2016). “Expressive Actions”, in C. Abell & J. Smith (eds.) *Emotional Expression: Philosophical, Psychological and Legal Perspectives*, Cambridge University Press, p. 73-94.
- Bennett, C. (2021). The Problem of Expressive Action, *Philosophy*, 96(2), 277-300.
- Berlin, I. (1958). *The Crooked Timber of Humanity*, Princeton: Princeton University Press.
- Betzler, M. (2007). Making Sense of Actions Expressing Emotions, *Dialectica*, 61(3), 447-466.
- Betzler, M. (2009). Expressive Actions, *Inquiry*, 52(3), 272-292.
- Cabrera, M. (2022a). ¿Qué es una emoción?, *Quaderns de filosofia*, 8(2), 145-91.
- Cabrera, M. (2022b). Duchenne Smiles are Actions not Mere Happenings: Lessons from the Debate on Expressive Action, *Theoria: An International Journal for Theory, History and Foundations of Science*, 37(2), 163-179.
- Cabrera, M. (in press). On the Intentionality of Shame and Pride, *Daimon Revista Internacional de Filosofía*.

- Cannon, W. B. (1927). The James-Lang Theory of Emotion: A Critical Examination and an Alternative Theory, *American Journal of Psychology*, 39: 106-124.
- Cobos, P. et al. (2002). Revisiting the James versus Cannon Debate on Emotion: Startle and Autonomic Modulation in Patients with Spinal Cord Injuries, *Biological Psychology* 61(3): 251-69.
- Corbí, J. E. (2012). *Morality, Self-knowledge and Human Suffering: An Essay on the Loss of Confidence in the World*. Routledge.
- Davidson, D. (1963). Actions, Reasons and Causes, *Journal of Philosophy*, 60(23), 685-700.
- Davidson, D. (1976). Hume's Cognitive Theory of Pride, *The Journal of Philosophy*, 73 (19): 744-757.
- De Sousa, R. (1987). *The Rationality of Emotion*, MIT Press.
- Deigh, J. (1994). Cognitivism in the Theory of Emotions, *Ethics*, 104: 824-54.
- Deigh, J. (2010). "Concepts of Emotions in Modern Philosophy and Psychology" en P. Goldie (ed.) *The Oxford Handbook of Philosophy of Emotion*, Oxford University Press, p. 17-40.
- Deonna, J. A., Rodogno, R. & Teroni, F. (2012). *In Defense of Shame: The Faces of an Emotion*, Oxford University Press.
- Deonna, J. A. y Teroni, F. (2012). *The Emotions: A Philosophical Introduction*, Routledge.
- Deonna, J. A. y Teroni, F. (2020). "Emotional Experience: Affective Consciousness And its role in Emotion Theory" in U. Kriegel (ed.) *Oxford Handbook of the Philosophy of Consciousness*, Oxford University Press.
- Dennett, D. C. (1995). "Do Animals Have Beliefs?", in H. Roitblat (ed.) *Comparative Approaches to Cognitive Sciences*, MIT Press.
- Döring, S. (2003). Explaining Action by Emotion, *The Philosophical Quarterly*, 53(211), 214-230.
- Dunn, R. (2006). *Values and the Reflective Point of View: On Expressivism, Self-knowledge and Agency*, Routledge.
- Frijda, N. H. (2007). *The Laws of Emotion*, Lawrence Erlbaum Associates Publishers.
- Goldie, P. (2000a). *The Emotions*, Oxford: Oxford University Press.
- Goldie, P. (2000b). Explaining expressions of emotion, *Mind*, 109, 25-38.
- Gordon, R. (1987). *The Structure of Emotions: Investigations in Cognitive Philosophy*, Cambridge, Cambridge University Press.
- Green, O. H. (1992). *The Emotions: A Philosophical Theory*, Dordrecht, Kluwer.

- Helm, B. (2001). *Emotional Reason: Deliberation, Motivation, and the Nature of Value*, Cambridge University Press.
- Helm, B. (2002). Felt Evaluations: A Theory of Pleasure and Pain, *American Philosophical Quarterly*, 39: 13-30.
- Helm, B. (2016). Emotional Expression, Commitment and Joint Value, In C. Abell & J. Smith (Eds.). *Emotional Expression: Philosophical, Psychological and Legal Perspectives*, Cambridge University Press, p. 95-114.
- Hursthouse, R. (1991). Arational Actions, *Journal of Philosophy*, 88, 57-68.
- James, W. (1884), What is an Emotion, *Mind*, 9: 188-205.
- Kim, J. (1993). *Supervenience and Mind*, Cambridge University Press
- McDowell, J. (1994). *Mind and World*, Harvard University Press.
- Mitchel, J. (2021). *Emotion as Feeling Towards Value: A Theory of Emotional Experience*, Oxford University Press.
- Moran, R. (2001). *Authority and Estrangement: An Essay on Self-knowledge*, Princeton University Press.
- Müller, M. J. y Wong, H. Y. (forthcoming). “Which Emotional Behaviors are Actions?”, in A. Scarantino (ed.) *Routledge Handbook of Emotion Theory*, London/New York: Routledge.
- Nussbaum, M. C. (2001). *Upheavals of Thought: The Intelligence of Emotions*, New York, Cambridge University Press.
- Perry, J. (1979). The Problem of the Essential Indexical, *Nous* 13, 3-21.
- Prinz, J., (2004). *Gut Reactions: A Perceptual Theory of Emotion*, New York, Oxford University Press.
- Scarantino, A. and Nielsen, M. (2015). Voodoo Dolls and Angry Lions: How Emotions Explain Arational Actions, *Philosophical Studies*, 172, 2975-2998.
- Pineda, D. (2019). *Sobre las emociones*, Madrid: Cátedra.
- Roberts, R. (2003). *Emotions: An Essay in Aid of Moral Psychology*, Cambridge: Cambridge University Press.
- Scarantino, A. (2010). Insights and Blindspots of the Cognitivist Theory of Emotions, *British Journal for the Philosophy of Science*, 61(4): 729–768.
- Smith, J. (2014). Are Emotions Embodied Evaluative Attitudes?, *Disputatio*, 6(38): 93-106.
- Smith, M. (1998). “The Possibility of Philosophy of Action”, in J. Bransen (ed.) *Human Action, Deliberation and Causation*, Dodrecht: Kluwer, p. 17-41.

- Solomon, R. (1993). *The Passions: Emotions and the Meaning of Life*, Hackett Publishing Company.
- Solomon, R. (2004). “Emotions, Thoughts, and Feelings: Emotions as Engagements with the World” in R. Solomon (ed.) *Thinking about Feeling: Contemporary Philosophers on Emotions*, Oxford/New York, Series in Affective Science, p. 76-88.
- Tappolet, C. (2016). *Emotions, Values and Agency*, Oxford: Oxford University Press.
- Tietjen, R. R. (2020). The Emotion of Self-reflexive Anxiety, *Journal of the American Philosophical Association*, 297-315.
- Teroni, F. (2016). Emotions, Me, Myself and I, *International Journal of Philosophical Studies*, 24: 433-451.
- Vendrell Ferran, I. (2018). “Emoción”, *Enciclopedia de la Sociedad Española de Filosofía Analítica* (URL: <http://www.sefaweb.es/emocion/>)
- Wiggins, D. (1987). *Needs, Values, Truth: Essays in the Philosophy of Value*, Oxford: Blackwell.
- Wollheim, R. (1999). *On the Emotions*, New Haven and London, Yale University Press.
- Zinck, A. (2008). Self-referential Emotions, *Consciousness and Cognition*, 17: 496-505.

Anexo 1: *On the Intentionality of Shame and Pride*

On the Intentionality of Shame and Pride¹

Sobre la intencionalidad de la vergüenza y el orgullo

Marta Cabrera²

Abstract

This paper argues against the widely held view according to which the main difference between shame and pride, and the rest of our emotions – such as fear, hate, surprise, joy or jealousy – is that the former are necessarily directed at the self as the intentional object of the emotion while the latter are typically oriented towards objects in the world other than the self. I examine three arguments in favour of this view and claim that further arguments should be provided in order to show that shame and pride are necessarily directed at oneself – *the doer* – rather than at an object different from the self – *the deed*.

Keywords: subject, object, evaluative properties, attention.

Resumen

Este artículo cuestiona la concepción ampliamente aceptada según la cual la principal diferencia entre la vergüenza y el orgullo, y el resto de nuestras emociones – miedo, odio, sorpresa, alegría o celos – es que las primeras se dirigen necesariamente al yo como objeto intencional de la emoción, mientras que las segundas se encuentran típicamente orientadas hacia objetos del mundo distintos del yo. Examino tres argumentos a favor de esta concepción y sostengo que el defensor de la misma debería aportar nuevos argumentos que muestren que la vergüenza y el orgullo se dirigen necesariamente a uno mismo – *el agente* – y no a un objeto distinto del yo – *el acto*.

Palabras clave: sujeto, objeto, propiedades evaluativas, atención.

* * * * *

¹ This work has been supported by the research projects PID2019-106420GA-I00 and PID2020-119588GB-I00 and by the grant BES-2017-081537 funded by MCIN/AEI/10.13039/501100011033 and “ESF Investing in your future”.

² PhD candidate at the Universitat de València. Her main areas of interest are philosophy of emotions, philosophy of mind and action theory, and she has recently published “Duchenne Smiles are Actions not Mere Happenings: Lessons from the Debate on Expressive Action” (2022) in *Theoria: An International Journal for Theory, History and Foundations of Science*, 37(2), 163-179 and “¿Qué es una emoción” (2022) in *Quaderns de filosofia*, 8(2), 145-91. Marta.Cabrera@uv.es

One of the most common assumptions among philosophers working on emotions is that there seems to be an important difference between shame and pride, and the rest of our emotions. According to a widely held view, shame and pride can be distinguished from the rest of our emotions by virtue of their specific intentionality, in particular *by their being necessarily oriented towards the subject of the emotion* (Helm 2001, Zinck 2008, Deonna and Teroni 2012, Deonna, Rodogno and Teroni 2012, Teroni 2016 and Tietjen 2020). This is to be contrasted with emotions such as fear, hate, surprise, jealousy, anger, joy or pity, which are typically oriented towards objects, people, states of affairs and events in the world other than the self. While this is a view that has become quite widespread among philosophers working on emotions, I will argue that the way of understanding the reflexivity of shame and pride it proposes seems problematic. Moreover, I will claim that such view is exclusively based on deviated forms of shame and pride, which prevents reflection on paradigmatic normative cases.

In section 1, I will introduce the view according to which shame and pride are necessarily directed at the subject of the emotion in contrast with the rest of our emotions. In section 2, I will examine “The Correctness Argument” and argue that the criteria that we use to assess our emotional responses do not allow us to establish that the subject rather than an object different from the subject is necessarily the intentional object of shame and pride. In section 3, I will review “The Phenomenological Argument” and argue that subject-focused cases of shame and pride are not a manifestation of their characteristic reflexivity but rather deviations from object-focused cases of such emotions. In section 4, I will address “The Grammatical Argument” and argue that feeling an emotion towards one’s features is not necessarily equivalent to feeling an emotion towards *oneself*. In section 5, I will conclude that further arguments should be provided in order to show that shame and pride are necessarily directed at oneself – *the doer* – rather than at an object different from the self – *the deed*. In addition, I will suggest that a plausible account of the reflexivity of shame and pride should be able to explain paradigmatic normative cases of such emotions, which are object-focused rather than subject-focused.

1. The Distinction

Philosophers working on emotions tend to think that there is an important distinction to be drawn between *reflexive* and *non-reflexive emotions*. “In the former class we find emotions such as guilt, shame, pride, embarrassment, and the like. The idea here is that

these emotions are distinctive in virtue of the fact that the subject has an attitude towards herself when undergoing the emotion” (Deonna and Teroni 2012, 18). While there is an ongoing debate regarding how best to account for the reflexivity of the different emotions that are usually classified as reflexive – for instance, on the basis of criteria such as responsibility, agency, accountability, self-ascription, etc. –, there is a broad consensus with respect to the idea that *shame* and *pride* are reflexive emotions in virtue of their *being necessarily directed at the self as the intentional object of the emotion*³. Although each of these emotions is related to the self from a different evaluative perspective – in shame, I see myself as unworthy or shameful; in pride, I see myself as praiseworthy – it would seem that, in either case, “the subject and intentional object of the emotion are identical” (Zinck 2008, 497. See also Deonna, Rodogno and Teroni 2012, Teroni 2016 and Tietjen 2020)⁴. While non-reflexive emotions – such as fear, hate, surprise, jealousy, anger, joy or pity – are typically oriented towards objects in the world other than the self, shame and pride would seem to be always directed at the subject who undergoes the emotion:

Non-reflexive felt evaluations are those emotions and desires whose targets are things in the world (...) By contrast, reflexive felt evaluations are those emotions and desires that focus on one’s being a certain kind of person (...) *pride*, in the sense I intend it here, is a positive evaluation of oneself or one’s character in light of one’s sense of the kind of person one ought to be. *Shame* or, more strongly, *mortification*, is a negative evaluation of oneself for not living up to one’s sense of who one ought to be (Helm 2001, 103-4, emphasis in the original).

³ This way of understanding the reflexivity of shame and pride is different from the way in which the reflexivity of guilt or remorse, for instance, is usually understood. Since the latter emotions cannot intelligibly be oriented towards oneself (I cannot feel guilt/remorse of or for myself), philosophers tend to account for their reflexivity in alternative ways. Since I will only be concerned here with the widespread view that explains the reflexivity of shame and pride in terms of the subject necessarily being the intentional object of such emotions, I will not get into the different ways in which we can understand the reflexivity of other emotions that are also thought to be reflexive.

⁴ Tietjen claims that “we may define self-reflexive emotions not just with reference to their material object [oneself] but also with regard to the concerns involved in them” because “theories that one-sidedly focus on the material object of self-reflexive emotions without reflecting on the concerns involved in them might be badly incomplete” (2020: 311). In any case, Tietjen assumes that, although this might not be the whole story, shame and pride are *directed at oneself as the intentional object of the emotion*. Since this is the idea that I will be questioning, it will be enough for our purposes that she agrees with the central claim of the view we will be discussing.

The idea here is that I am the intentional object of shame not just when I am ashamed of my big nose, for instance, but also when I feel ashamed due to the lie I have told a friend of mine. The thing I am *ashamed of*, in both cases, is *myself*. Likewise, I am the intentional object of pride not only when I am proud of being a good mother to my children, for instance, but also when I proudly contemplate my business. The thing I am proud of, that is, the object that I represent as *something worthy of pride* is, in both cases, *myself*. This is to be contrasted, as we said, with emotions such as fear, hate, surprise, jealousy, anger, joy, or pity, which are typically directed at objects that are different from the self. The shared thought seems to be that the self is not distinctively implicated in these emotions, or, in other words, that such emotions “do not critically involve the self” (Tracy and Robins 2007, 7).

One could certainly reply that some of these emotions can also be reflexive on some occasions: I may fear *the dog* in front of me (non-reflexive fear) but I may also fear *myself* because of my impulsive behaviour (reflexive fear), I may hate *my annoying neighbour* (non-reflexive hatred) but I may also hate *myself* for how insecure I am (reflexive hatred), I may be surprised by *a piece of news* (non-reflexive surprise) but I may also be surprised by *my ability* to quickly pick up a language (reflexive surprise), etc. However, as advocates of the reflexive/non-reflexive distinction claim, this sort of reflexiveness – in which the self or one of its features is the intentional object of a typically non-reflexive emotion – seems to be of a *contingent* kind: it is constitutive of these emotions that they *can* have the self as their intentional object, but in the case of shame and pride, the self is the *only* intentional object that they can have. Emotions like fear, hate or surprise are not necessarily about the self: they tend to be directed at people, other living creatures, physical objects, events or states of affairs that surround us – in which case, they would seem to free us from preoccupation with the self⁵.

⁵ One could also reply that, just as fear, hate or surprise can be directed at the self or at an object different from the self, shame and pride might also be oriented towards the self on some occasions – as when I am ashamed of my nose or proud of being a good mother – while on others, such emotions might be oriented towards objects that are *different from the self*. Perhaps, the thought goes, I can be ashamed of *the lie* I have told my friend or proud of *the business* I have built, and not necessarily ashamed or proud of myself. Although we can intuitively accept that shame and pride are directed at myself when I am ashamed of my nose or proud of being a good mother (the same way my reflexive fear or reflexive hatred are oriented towards myself), it is not obvious that this is so when the shame I feel has to do with the lie I have told my friend or the pride I feel has to do with the business I have built. That is, it does not seem obvious that, in the latter cases, *what I am ashamed of or proud of is myself*. This is roughly the view I will argue for through the arguments that I develop in the following sections, against the idea that shame and pride are necessarily directed at the self.

In what follows, I briefly review the need for some arguments in favour of this idea (Teroni 2016) and argue that the suggested way of distinguishing shame and pride from the rest of our emotions is subject to some important problems. If the objections I develop in the coming sections are plausible, the idea that the reflexivity of shame and pride has to do with the fact that they are necessarily directed at oneself will be in need of further justification.

2.The Correctness Argument (CA)

According to Teroni (2016), shame and pride are necessarily directed at the subject of the emotion because, in their case, the subject *must* exemplify the relevant evaluative property – shamefulness or praiseworthiness – in order for the emotion to be correct. We will hereby refer to such argument as the Correctness Argument (CA)⁶.

A reflexive emotion takes the subject who undergoes it as its particular object, and is correct if and only if the subject exemplifies the evaluative property that constitutes the formal object of the emotion. A non-reflexive emotion does not abide by the same requirement: its correctness conditions require that a particular object, distinct from the subject who undergoes it, exemplify the relevant formal object (Teroni 2016, 5)⁷.

CA seems to be based on the idea that the identification of the object of an emotion is linked to the criteria of correctness that allow us to evaluate an emotional response as correct or incorrect. According to this idea, the object of an emotion is what sets the criterion for evaluating the subject's emotional response: my fear of the dark is a correct or proportional response if *the dark* merits fear, therefore, the dark is the object of my fear. In more standard terms, the object of the emotion is what exemplifies (or not, if the emotion is inadequate) the evaluative property that corresponds to the type of emotion in question. We can therefore reformulate CA as follows: *if the criterion of correctness of an emotion is provided by the subject and her features, we are dealing with a reflexive*

⁶ The labels “The Correctness Argument” (CA) (s. 2.), “The Phenomenological Argument” (PA) (s. 3.) and “The Grammatical Argument” (GA) (s. 4.) are mine.

⁷ “Formal objects” of emotions correspond to the evaluative properties that allow us to distinguish between different types of emotion: dangerousness is the formal object of fear, unworthiness or shamefulness is the formal object of shame, loss is the formal object of sadness, praiseworthiness is the formal object of pride, etc. (Kenny 1963).

*emotion; the correctness criteria for shame and pride are provided by the subject and her features, and thus they are reflexive emotions*⁸.

It seems clear that CA's treatment of reflexive emotions easily applies to those cases where we have already assumed that the subject's emotion is directed at herself: my shame for my big nose is correct or proportional *if my nose is big* (which is a shameful thing) and the pride I feel for being a good mother is correct *if I am, in fact, a good mother* (which is something praiseworthy). Similarly, my fear of my impulsive behaviour is correct *if I am an impulsive person* (which makes me dangerous), my hatred of my insecurities is correct *if I am an insecure person* (which is something hateful), and so on⁹. In other words, in all these cases, it does seem that we evaluate the subject's emotional response by looking at the subject herself and her features – the emotion “is correct if and only if the subject exemplifies the evaluative property that constitutes the formal object of the emotion” (Teroni 2016, 5). This means that, in all these cases, the subject is the intentional object of the emotion: she is ashamed of herself, proud of herself, scared of herself and hates herself, respectively. So far, CA seems convincing.

What happens, though, when what I am ashamed of is the lie I have told my friend or what I am proud of is my business? Teroni seems to think that, in these cases, the correctness of the emotional response is also tied to the subject's exemplifying the relevant evaluative property. That is, my shame is correct if I am a shameful person due to the lie I have told my friend and my pride is correct if I am a praiseworthy person for how I have made my business prosper. Accordingly, since I exemplify the evaluative properties that are required in order for my shame and my pride to be correct, *I am the intentional object* of both emotions. On the other hand, and this is where non-reflexive emotions differ from shame and pride according to Teroni, the fear I feel towards the dog in front of me and the hatred that I experience towards my annoying neighbour are correct emotions if, respectively, the dog is dangerous and my neighbour is a hatefully annoying person. Since the features of the dog and the neighbour are what allow us to evaluate the correctness or proportionality of my emotional responses, we can conclude that the dog and my neighbour are the intentional objects of my fear and hatred.

⁸ I thank an anonymous referee for pressing me to formulate this argument more clearly.

⁹ I do not agree with the idea that feeling an emotion towards certain features or aspects of oneself implies that *the subject herself* is the intentional object of the emotion. For the sake of the argument, however, we will provisionally assume that this is so. This issue will be addressed in section 4.

This apparently clear and easy way of distinguishing shame and pride from the rest of our emotions is subject, however, to some difficulties. In order to examine these difficulties, we should first note that when assessing the situation to which an emotion responds (in order to determine whether the subject's emotional response is adequate or not) such a situation *can be described in many ways* and it is not always necessary to make *explicit* reference to the evaluative properties of the object of the emotion. For instance, we can say that my fear of the dog is correct if *I am in danger* or my hatred of my annoying neighbour is correct if *I have to put up with his conversations every day*. In each case, what I analyse in order to see whether the emotion in question is correct is *the subject* of the emotion together with certain aspects related to her. This move, however, *does not imply that the subject is the object of the emotion*, that is, it does not imply that we are emotionally responding to ourselves: although the features of the subject are what allow us to evaluate the emotional response (if I am not in danger, my fear is not a good response), this is only possible because such features *implicitly* refer to the features of the intentional object of the emotion (I am not in danger if the dog is not dangerous).

I believe Teroni would agree that, although I can assess whether my fear is correct by observing whether I am in danger or whether my hatred is correct by observing whether I have to put up with my neighbour every day, the relevant features for assessing such emotional responses are, ultimately, whether *the dog is dangerous* and whether *my neighbour is annoying*. After all, it is the dog and my neighbour that must exemplify the relevant evaluative property for my emotions to be correct. And this is so because *the dog and my neighbour are the intentional object* of my fear and hatred.

Now that we have clarified that the subject's features can implicitly refer to the features of the intentional object and that this does not mean that the subject herself is the object to which her emotion responds, we can look into the difficulties that affect CA. Let us go back, then, to the case of shame and imagine that, out of self-interest or some deep insecurity, I have given a friend of mine some false information that will make it extremely hard for her to pass an important exam. When I think, after a while, about the lie I have told her, I am overcome with shame: not only it is wrong to lie to a friend, but I am significantly harming her by not giving her the right information. From my perspective, what I have done to her is a serious thing. What are the relevant features that allow us to determine whether my feeling shame in these circumstances is correct or appropriate? It seems clear that shame is a response to shamefulness or unworthiness. The question is: *what is the shameful or unworthy object to which I am responding* and that

allows us to assess the proportionality of my emotional response? As we have seen, according to Teroni, this object is myself: shame is a response to my own shamefulness. However, it seems that, while we can look at the subject and her features (as in the case of the dog or my neighbour), we can also look at *the lie itself*. Is not the lie, after all, something shameful and unworthy in the face of which it is appropriate to feel shame? Perhaps, just as we sometimes assess whether my fear of the dog is appropriate in terms of whether I am in danger because we implicitly refer to the dangerousness of the dog, when we assess whether the shame I feel is appropriate in terms of whether I am a shameful person, we are also implicitly referring to the shamefulness of the lie.

So, in the case of shame, what is shameful – the lie or myself? And, in the case of pride, what is praiseworthy – my business or myself? A natural answer to these questions is: *both*. This is an important difference with respect to the case of fear: standing before something dangerous puts me in danger, but it does not make me a dangerous person. By contrast, doing something shameful or praiseworthy makes me a shameful or praiseworthy person. In a way, the evaluative property is “transferred” to the agent of the action. Thus, in the case of shame and pride, we can say that both *the act* (that is, the *deed* – lying and building a business) and *the subject* (the *doer*) exemplify the evaluative property in question. But *does it follow from this that the subject is the intentional object of such emotions* (since she exemplifies the evaluative property that is required for her emotion to be correct)? Does it follow from this that the subject is *always* responding to herself when feeling shame and not to something other than herself? Whoever considers that, from what we have said, it does indeed follow that the subject is the intentional object of shame (and not the lie) would have to provide some arguments in favour of this idea, for it seems that there are *two candidates* for the object of the emotion, since both the shameful lie (the *deed*) and the shameful subject (the *doer*) exemplify the relevant evaluative property.

Given the two candidates for the object of the emotion, one might point out that the lie has an important *priority* over the liar: the subject is shameful insofar as she has performed a shameful act, but not the other way around. What ultimately determines whether the subject is a shameful person is the lie she has told. So, even if we want to assess the correctness or proportionality of shame in terms of whether the subject is a shameful person, we will inevitably have to refer to the lie and its shamefulness. It would seem, therefore, that the object that ultimately provides us with the criterion for evaluating the emotional response is the lie. We might conclude, then, that *the lie*, and not the subject,

is the intentional object of shame. Focusing on the features of the subject seems rather (as in the previous examples of fear and hatred) an implicit way of referring to the features of the lie.

The ideas presented so far, rather than definitively determining what the object of reflexive emotions are, suggest that the subject/object division presupposed in this discussion is quite problematic. The idea that the evaluative property that corresponds to the emotion we have to assess has to be exemplified by the subject *or* by an object other than the subject rests on a rather artificial way of articulating the situation to which the subject's emotion responds. As we are seeing through the examples we are examining, evaluative properties seem to depend both on the subject and on objects other than the subject. The sort of things we can say about this issue, however, will depend very much on the conception of evaluative properties to which we are committed. This question is part of a long and complex debate which we will not go into here¹⁰.

3. The Phenomenological Argument (PA)

The following argument turns away from issues of correctness and focuses on the phenomenological differences that may exist between shame and pride, and non-reflexive emotions, for which reason we will refer to such argument as the Phenomenological Argument (PA).

The characteristic reflexivity of emotions such as shame and pride, which is manifested in the attention directed at oneself and one's features, is here [amusement, admiration, joy, pity, gratitude and compassion] replaced by an attention wholly directed outside the self (...) [a]ll these emotions in their normal activities free us from preoccupation with self (Teroni 2016, 6)¹¹

This idea is very different from the one expressed by CA. In the previous argument, the aim was to show that, in the case of shame and pride, the subject is the object of the emotion because the subject and her features are what provide the criteria of correctness

¹⁰ For the debate on the relation between emotions and evaluative properties see Wiggins (1987), Goldie (2000), Dunn (2006), Corbí (2012), Deonna and Teroni (2012) and Tappolet (2016). A related issue we will not go into either is the asymmetry that appears to exist between the third and the first person in the determination of the shameful or blameworthiness of a deed. Talk of mere correctness – at least in the terms we have been discussing so far – ignores the fact that someone's shame or guilt may be a rational, correct and proportional response to a situation even if, from a third person perspective, we do not think that the subject has done anything shameful or blameworthy. Cases of survivor's guilt or shame are especially telling in this respect (Corbí 2012).

¹¹ With the last sentence, Teroni is quoting Shand (1918, 227).

of the emotion (as opposed to what happens with non-reflexive emotions). With PA, on the other hand, the reasoning seems to be as follows: *if the subject focuses her attention on herself and her features when experiencing an emotion, we are dealing with a reflexive emotion; the subject's attention is directed towards herself and her features when she feels shame and pride, and thus they are reflexive emotions.* By contrast, when the subject experiences amusement, joy or pity, her attention is “wholly directed outside the self” (2016, 6), thus they are non-reflexive emotions.

The idea here is that the fact that the subject tends to focus her attention on herself – and, more precisely, on how she looks from the point of view of others – when experiencing shame and pride is indicative that she herself is the object of her emotion in such cases. This seems to be Teroni’s suggestion when he claims that attention towards oneself is *a manifestation* of “the characteristic reflexivity of emotions such as shame and pride” (2016, 6). This claim is based on the assumption that when I feel ashamed of something, my thoughts and feelings revolve around me – “why did I do this?”, “I am a terrible person”, “I should have known better”, etc. – while when experiencing fear of the dog or hatred towards my neighbour my attention is typically focused on salient features of the dog (its teeth, claws, etc.) or my neighbour (the way he speaks, how repetitive he is, etc.). However, is it really the case that shame and pride are characteristically subject-focused in the way suggested by Teroni?

Contrary to what is presupposed by PA, it seems that those cases in which the subject of shame or pride focuses on herself and on how she might look from the point of view of others constitute *a very particular kind of case* that appears to be a *deviation from the normatively paradigmatic case*. In order to illustrate this idea, Moran (2001) introduces the example of Fred Vincy, one of the main characters of Eliot’s novel *Middlemarch*, who feels ashamed after having lost the money he had borrowed from the Garths. In the example Moran examines, Fred feels sorry for what he has done but, at some point, all he seems concerned with is *how he looks from the point of view of the Garths*, rather than with the harm caused to them. As Moran points out, “[h]is attention to how he is to be esteemed interferes with reflection on the significance of what he did” (Moran 2001, 189).

Curiously enough, his pain in the affair beforehand had consisted almost entirely in the sense that he must seem dishonourable, and sink in the opinion of the Garths: he had not occupied himself with the inconvenience and possible

injury that his breach might occasion them, for this exercise of the imagination on other people's needs is not common with hopeful young gentlemen. Indeed we are most of us brought up in the notion that the highest motive for not doing a wrong is something irrespective of the beings who would suffer the wrong (Eliot 1871, Chapter 24, quoted in Moran 2001, 189).

The point of this example is to show that by focusing his attention on himself, Fred is neglecting, in an important sense, the object of his emotion¹².

According to this, there seems to be two possible configurations of shame: on the one hand, we have *subject-focused shame* – where the subject focuses on how she looks from the outside – and, on the other, *object-focused shame* – where the subject focuses on what she has done and on the harm caused to others. The former constitutes the kind of shame we tend to *censor* as an inadequate, narcissistic, insensitive or even fearful response on the part of the subject: as Eliot's and Moran's fragments show, this does not seem to be the way in which one *should* respond when one has harmed someone, *even if*, in many cases, this is how we *normally* respond. The normative requirement arises from the fact that, by focusing exclusively on herself, the subject ceases to acknowledge the significance of what she has done and distances herself from the harm and the needs of those affected by her actions. This seems to be the reason why refocusing one's attention away from oneself and towards the harm caused is seen by us as an ethical achievement (Corbí 2012). Object-focused experiences of shame are hence the normatively paradigmatic cases that allow us to criticise subject-focused cases as deviations.

Turning back to our previous examples, this means that when I feel ashamed of the lie I have told my friend, I can focus on myself, on what a bad friend I am, on what others will say of me, etc. (subject-focused shame) or, in contrast, I can focus on my friend, on how much I have hurt her, on what she may need from me now, etc. (object-

¹² It should be noted that Moran originally analyses such an example as a case of guilt. However, Eliot's words are that Fred's worry is "that he must seem *dishonourable*, and sink in the opinion of the Garths" (Eliot 1871, Chapter 24, quoted in Moran 2001, 189, emphasis added). Losing one's honour is something for which one feels shame, not guilt, so the reading I suggest seems to be plausible given the way Eliot describes Fred's thoughts and feelings. Although for reasons of space I will not be able to dwell on this issue, I acknowledge that more should be said about the difference between shame and guilt since – according to my arguments – this difference cannot be that the former is intentionally directed at the self while the latter is intentionally directed at an object different from the self. As I said in section 1, there is an ongoing debate regarding the criteria that allow us to classify an emotion as reflexive (responsibility, agency, accountability, self-ascription, etc.). Perhaps, the difference between shame and guilt lies in the role that accountability plays within these experiences, or in the kind of failure to which the agent is responding in each case. Although more should clearly be said about this issue, all I can argue for at this point is the negative conclusion that shame cannot be distinguished from guilt in virtue of the subject being the intentional object of the emotion.

focused shame). The same goes for pride: when I am proud of my business, I can focus on myself, on how great my achievements are, on how many people must admire me, etc. (subject-focused pride) or, in contrast, I can focus on the business itself, on how smoothly things work, on how good the atmosphere among my colleagues is, etc. (object-focused pride). We might even think that genuine pride is “selfless”¹³.

From all this, it seems that we can conclude that, contrary to PA, cases in which the subject focuses her attention on herself and on how she looks are of a very specific kind, which we censure as inappropriate. There seem to be more proportional and adequate forms of shame and pride in which the subject focuses on the shameful or praiseworthiness of her deed rather than on herself. This brings into question the idea that attention towards oneself is *a manifestation* of “the characteristic reflexivity of emotions such as shame and pride” (Teroni 2016, 6). Such a claim ignores what we have just called proportional and adequate forms of shame and pride and focuses exclusively on their deviations. This is an important distinction that we should keep in mind given the fundamental role it plays in the everyday practices through which we regulate our emotions. Additionally, we can see how CA turns against PA in this respect: the shameful or praiseworthiness of a deed determine that when the subject focuses on herself, she is turning away from the object of her emotion¹⁴.

Moreover, the idea that the subject’s attention is “wholly directed outside the self” (Teroni 2016, 6) when feeling typically non-reflexive emotions, such as amusement, admiration, joy, pity, gratitude and compassion does not seem to be correct either. When undergoing these kinds of emotions the subject can also shift her focus of attention and focus on herself (Dunn 2006), something that also allows us to distinguish between proportional or disproportional, adequate or inadequate, narcissistic and non-narcissistic

¹³ There may be many other configurations beyond completely subject-focused and completely object-focused shame and pride (Goldie 2000, 17). Our emotions are closely related to our cares and concerns, so it is only natural that when experiencing an emotion our attention and thoughts occasionally shift back and forth towards ourselves (see Dunn 2006 and Corbí 2016).

¹⁴ Someone may, nevertheless, put forward the following counterexample to the idea that cases of subject-focused shame are normatively inappropriate: noticing that I am a bad person, I may feel shame and worry about my character in general, which would not only be an appropriate response but even a virtuous one, since I would be concerned with improving myself as a person (I thank an anonymous referee for suggesting this counterexample to me). My response to this objection consists in pointing out that such an emotional response can only be a morally virtuous one if my concern is connected to the harm caused to others – to improving myself so that *I will not harm them again* – and not merely with whether I can be considered a good person. I think that, when the former is the case, my shame cannot be seen as completely subject-focused for the others are at the center of my concern: my feelings and my wish to improve is not strictly about me, but rather about them. In any case, these considerations seem to bring us back to the problem of drawing a clear line between the subject and the object of an emotion.

forms of these emotions. This is something that Teroni himself seems to admit when he introduces the case of someone who pities another because feeling such emotion triggers a positive image of herself, and not because she is worried about the other person's situation. This kind of case, says Teroni, must be explained in terms of narcissistic personality trait (2016, 7).

To summarise, the assumption that the subject's attention is directed towards herself whenever she experiences shame and pride does not seem to be correct. As mentioned, when this occurs, we are dealing with very specific forms of these emotions. In any case, even if there is a psychological tendency – which is normatively deviated – to focus on one's own image when experiencing shame and pride, *this does not imply that the subject is the intentional object of such emotions*¹⁵.

4. The Grammatical Argument (GA)

The last argument – which we will refer to as the Grammatical Argument (GA) – focuses on the way we speak about our emotions and the changes in meaning that can take place when we introduce explicit reflexive components such as “herself”, “myself”, etc.

We commonly report pride or shame episodes by saying such thing as ‘Mary is proud of the canvas’ or ‘Mark is ashamed of his big nose’, but these constructions can easily be completed in the following fashion: ‘Mary is proud of *herself* because she painted this canvas’, ‘Mark is ashamed of *himself* because of his big nose’. This rule does not apply to non-reflexive emotions. (Teroni 2016, 5, emphasis in the original).

According to GA, then, *if we can insert an explicit reflexive component in sentences that refer to emotions without changing the meaning of the sentence, we are dealing with a reflexive emotion; we can insert an explicit reflexive component in sentences that refer to*

¹⁵ The strategy that I have developed in this section can also be applied to another possible argument in favour of the idea that shame and pride are necessarily directed towards the self (I thank an anonymous referee for pointing out such argument to me). One could think that the fact that shame and pride involve action tendencies that are directed towards oneself shows that such emotions necessarily take the self as their intentional object – for instance, hiding from the world or exhibiting submissive behaviour in the case of shame, or standing up tall in the case of pride. These cases, however, seem to be *subject-focused* examples of shame and pride (although it is not clear to me why exhibiting submissive behaviour should be thought of as being directed towards oneself). In more object-focused cases, our action tendencies might be completely different: trying to make up for what we have done in the case of shame or displaying our work so that everyone can enjoy it in the case of pride, for instance. In any case, and as I have concluded in section 3, even if there is a tendency to focus on oneself when experiencing shame and pride, this does not imply that the subject is the intentional object of such emotions.

shame and pride without changing their meaning – all we do is complete the information already given –, and thus they are reflexive emotions. In the case of sentences that refer to non-reflexive emotions, by contrast, the insertion of a reflexive component – if grammatical at all – *does* change the meaning of the sentence. As Teroni points out, getting upset at a comment that someone has made is very different from getting upset *at oneself* because one's naiveté is the reason why the other person made such comment (Teroni 2016, 7).

According to our discussion so far, the case in which Mark is ashamed of his big nose is different, in a relevant sense, from the case in which Mary is proud of the canvas she has painted. In the former case, we have assumed that the intentional object of shame is the subject himself, whereas, in the latter (which is equivalent to the case in which I am proud of my business), we have pointed out that it is not clear that the subject is the intentional object of pride (unlike when I am proud of being a good mother – that is, of one of my features). Someone might say that, for this reason, it is not clear that GA applies to the case where Mary is proud of her canvas or to the case where I am proud of my business: it is not clear that what we are proud of is ourselves (for the reasons given in section 2.)

Therefore, there does seem to be an important difference between saying that I am proud of my canvas or my business, on the one hand, and saying that I am proud of *myself* because I painted this canvas or because I made this business prosper, on the other. In the latter cases, rather than the canvas or the business, what I am proud of is some trait or ability of mine that I consider praiseworthy. In such cases, intuitively, the subject does seem to be the intentional object of pride¹⁶. In the former case, however, the object of pride seems to be rather the canvas or the business. This becomes more obvious if we notice that we can be proud of our canvas or our business without necessarily being proud of ourselves.

For these reasons, one might think that the transformation suggested by Teroni does not work in the case where Mary is proud of the canvas she has painted. However, as we said, the case in which Mark is ashamed of his big nose is different: since his shame is directed towards himself and his features, it seems that we can assume that Mark is ashamed of *himself* because of his big nose.

¹⁶ As mentioned in note 9, we have provisionally accepted, for the sake of the argument, that feeling an emotion towards certain features of oneself implies that *the subject herself* is the intentional object of the emotion.

Thus far in the paper, we have assumed that experiencing an emotion towards one's own features is equivalent to experiencing that emotion towards oneself. We can now see, however, that just as the subject/object distinction was problematic when we talked about the criteria of correctness of emotions, neither can we draw a clear line between features that belong exclusively to the subject and features that are related to objects other than the subject. Can we really transform "Mark is ashamed of his big nose" into "Mark is ashamed *of himself* because of his big nose"? Just as we can be proud of our canvas or our business without necessarily being proud of ourselves, it seems that we can be ashamed of our nose without necessarily being ashamed of ourselves: being ashamed of oneself seems much deeper than feeling shame towards one of our features¹⁷.

This idea, moreover, seems to apply also to the reflexive cases of fear and hatred. According to Teroni, we should be able to transform "I fear my impulsive behaviour" into "I fear *myself* for my impulsive behaviour" and "I hate my insecurities" into "I hate *myself* for my insecurities" with no relevant changes. But, again, it seems that we can fear our impulsive behaviour without necessarily fearing ourselves, and we can hate our insecurities without necessarily hating ourselves – fearing or hating oneself seem much deeper than feeling fear or hatred towards one of our features.

If this is the case, it seems that we would have to restrict much more what it takes for an emotion to be directed towards oneself, and this is a question that requires a far more extensive and deep analysis than the one we intend to carry out here. The conclusion we draw from the analysis of this last section is that we do not seem to be able to make the transformations proposed by GA without relevant changes related to the intentional object of shame and pride.

¹⁷ This seems to be evident if we think about the different impact that being ashamed of a physical trait one has and dislikes, for instance, and being ashamed of oneself may have on someone's life. In the first case, the person's shame might not affect her life in very significant ways: it might just be expressed in the way she tends to hide the part of her body she is ashamed of with her clothes, for instance, but, otherwise, it might not affect her overall behaviour and concerns. More importantly, she might not necessarily be ashamed of herself, that is, of who she is. In the second case, however, the impact of her shame on her life might be of a more substantial kind, determining her personality and behaviour in such ways as to become an important obstacle for her. Although, clearly, much more should be said about this contrast – it might be interesting, for instance, to reflect on those cases in which, due to one's identification with certain physical standards, one's shame at certain physical features *does involve* shame at oneself – it is enough for our purposes that feeling shame at oneself does not seem to be necessarily equivalent to feeling shame at one's features.

5. Conclusions

I have tried to show that the proposed arguments do not seem to provide sufficient grounds for considering that shame and pride are always intentionally directed at the self while the rest of our emotions are oriented towards objects in the world other than the self. Firstly, regarding CA, we have seen that there are always two candidates for the object of shame and pride – the shameful/praiseworthy deed and the shameful/praiseworthy doer – and that the shamefulness/praiseworthiness of the deed seems to have an important priority over the shamefulness/praiseworthiness of the doer in determining the correctness or proportionality of the subject's emotional response. Secondly, with respect to PA, we have argued that the object of an emotion need not be the focus of the subject's attention during the emotional experience and that, when this is the case, we are faced with a very particular kind of shame or pride that appears to be a deviation of the normatively paradigmatic case. Lastly, in relation to GA, we have seen that the insertion of an explicit reflexive component in sentences that refer to shame or pride does seem to change the meaning of such sentences. Therefore, it seems that if someone wants to hold on to the idea that the reflexivity of shame and pride is to be found at the level of their intentional object – in particular, in their being necessarily directed at the subject of the emotion – they should provide new arguments in favour of this idea.

More importantly, our discussion suggests that contemporary philosophical research on shame and pride significantly ignores what we have called paradigmatic normative cases, that is, cases of shame and pride in which the subject focuses on the object of her emotion rather than on herself. These cases clearly exist and, as was said before, in certain contexts, they can be seen as ethical achievements (Moran 2001 and Corbí 2012). The everyday practices through which we regulate our emotions show that object-focused cases of shame and pride are especially relevant to us, for we criticise people who fearfully or narcissistically focus on their own image and neglect the shameful or praiseworthy object to which their emotion might be responding. Ignoring such cases, as the widespread view we have been examining here does, however, has led to a view of the reflexivity of shame and pride that, for all the reasons given above, seems quite implausible. The implausibility of such view, though, should not be taken to entail that shame and pride should ultimately be considered analogous to non-reflexive emotions. All I claim is that we seem to be in need of an alternative account of the reflexivity of shame and pride that does not claim that such emotions are *necessarily* directed at the self

as the intentional object of the emotion and that is able to explain their object-focused cases¹⁸.

Acknowledgments

I want to thank Josep Corbí and Jordi Valor for their valuable comments and suggestions as well as two anonymous referees, Jonathan Mitchell and Anna Giustina for their helpful insights on earlier versions of this paper. I am also indebted to the audiences in the *Valencia Colloquium in Philosophy* (Universitat de València) and the *Perspectives on Intentionality and Consciousness Conference* (University of Waikato).

References

- Clark, J. A. (2013). Integrating basic and higher-cognitive emotions within a common evolutionary framework: Lessons from the transformation of primate dominance into human pride. *Philosophical Psychology*, 26(3), 437–460.
- Corbí, J. E. (2012). *Morality, self-knowledge and human suffering: an essay on the loss of confidence in the world*. Routledge.
- Corbí, J. E. (2016). Self-knowledge, deliberation, and the fruit of Satan. *Acta Analytica*, 32(2), 245–26.
- Deonna, J. A., Rodogno, R. & Teroni, F. (2012). *In defense of shame: the faces of an emotion*. Oxford University Press.
- Deonna, J. A. & Teroni, F. (2012). *The emotions: a philosophical introduction*. Routledge.
- Dunn, R. (2006). *Values and the reflective point of view: on expressivism, self-knowledge and agency*. Routledge.
- Ekman, P. (1999). Basic emotions. In T. Dalgleish & T. Power (Eds.), *The handbook of cognition and emotion* (pp. 45–60). Sussex, UK: John Wiley & Sons, Ltd.

¹⁸ These conclusions may also have some bearing on the wider field of philosophy of emotion, for instance, on the debate regarding the continuity among the category of emotion. Authors like Ekman (1991), Griffiths (1997) or Tracy, Robins and Tagney (2007) insist that our emotions must be divided into two different categories: basic emotions and higher-cognitive emotions. Among the latter, they argue, we find the subclass of self-conscious emotions – such as shame and pride – which require cognitive capacities that allow for self-awareness, self-evaluation and awareness of other selves, unlike so-called basic emotions. The idea that shame and pride are not necessarily directed at the self as the intentional object of the emotion – that is, that they are responses to objects different from the self – might contribute to the alternative view that there is a continuity among all our emotions (see Roberts 2009, Clark 2013 and Fugate 2015).

- Eliot, G. (1871). *Middlemarch*. Wordsworth Classics.
- Fugate, J. M. B. (2015). Evolutionary Constraints and Cognitive Mechanisms in the Construction of an Emotion. En L. F. Barrett & J. A. Russell (Eds.), *The Psychological Construction of Emotion* (pp. 399–426). The Guilford Press.
- Goldie, P. (2000). *The emotions*. Oxford: Oxford University Press.
- Griffiths, P.E. (1997). *What emotions really are: The problem of psychological categories*. Chicago: University of Chicago Press.
- Helm, B. (2001). *Emotional reason: deliberation, motivation, and the nature of value*. Cambridge University Press.
- Kenny, A. (1963). *Action, emotion and will*. Routledge.
- Moran, R. (2001). *Authority and estrangement: an essay on self-knowledge*. Princeton University Press.
- Roberts, R. C. (2009). The sophistication of non-human emotions. En R. W. Lurz (Ed.), *The Philosophy of Animal Minds* (pp. 218–236). Cambridge University Press.
- Shand, A. (1918). Emotion and Value. *Proceedings of the Aristotelian Society*, 19, 208–235.
- Tappolet, C. (2016). *Emotions, values and agency*. Oxford: Oxford University Press.
- Tietjen, R. R. (2020). The emotion of self-reflexive anxiety. *Journal of the American Philosophical Association*, 297-315.
- Teroni, F. (2016). Emotions, me, myself and I. *International Journal of Philosophical Studies*, 24, 433-451.
- Tracy, J. L., Robins, R. W. & Tagney, J. P. (2007). *The self-conscious emotions: theory and research*. The Guildford Press.
- Tracy, J. L. & Robins, R.W. (2007). The self in self-conscious emotions: a cognitive appraisal approach. In J. L. Tracy, R. W. Robins & J. P. Tagney (Eds.), *The self-conscious emotions: theory and research* (pp. 3–20). The Guildford Press.
- Wiggins, D. (1987). *Needs, values, truth: essays in the philosophy of value*. Oxford: Blackwell.
- Zinck, A. (2008). Self-referential emotions. *Consciousness and Cognition*, 17, 496-505.

**Anexo 2: Duchenne Smiles are Actions not Mere Happenings:
Lessons from the Debate on Expressive Action**



Duchenne smiles are actions not mere happenings: lessons from the debate on expressive action

(*La sonrisa de Duchenne es una acción no un mero suceso:
lecciones del debate sobre las acciones expresivas*)

Marta CABRERA*

Universitat de València

ABSTRACT: In this paper, I will argue that, contrary to what is generally assumed in the debate on expressive action, we do not have good reasons to exclude facial and bodily expressions of emotion such as smiling or frowning from the category of actions. For this purpose, I will compare facial and bodily expressions of emotion with simple expressive actions, such as jumping for joy or covering one's face in shame. I will try to show that simple expressive actions cannot be presented as actions while excluding facial and bodily expressions of emotion from this condition. My contention will then be that either both sorts of behaviour are to be identified as actions or neither is. The latter sounds rather implausible, though, as we would have to assimilate jumping for joy or covering one's face in shame to spasms, which conflicts with the way we relate to such behaviours. My conclusion will then be that both simple expressive actions and facial and bodily expressions of emotion should be included within the category of actions, at least on the basis of the main assumptions in the current debate on expressive action.

KEYWORDS: rationality; instrumentality; voluntariness; responsiveness; evaluative perspective.

RESUMEN: En este artículo argumentaré que, a diferencia de lo que suele asumirse en el debate sobre las acciones expresivas, no tenemos buenas razones para excluir a las expresiones faciales y corporales de las emociones, como sonreír o fruncir el ceño, de la categoría de acciones. Para ello, compararé las expresiones faciales y corporales de las emociones con acciones expresivas simples como, por ejemplo, saltar de alegría o cubrirse avergonzado la cara. Intentaré mostrar que no puede presentarse a las acciones expresivas simples como acciones y excluir, al mismo tiempo, a las expresiones faciales y corporales de las emociones de esta categoría. Sostendré, por tanto, que, o bien ambos tipos de comportamiento son acciones, o bien ninguno lo es. Esto último, no obstante, resulta bastante problemático, ya que tendríamos que asimilar el saltar de alegría o el cubrirse avergonzado la cara a meros espasmos, lo cual entra en conflicto con la manera en la que nos relacionamos con este tipo de comportamientos. Mi conclusión será entonces que tanto las acciones expresivas simples como las expresiones faciales y corporales de las emociones han de ser incluidas en la categoría de acciones, al menos, dados los presupuestos principales del debate actual sobre acciones expresivas.

PALABRAS CLAVE: racionalidad; instrumentalidad; voluntariedad; capacidad de respuesta; perspectiva evaluativa.

* **Correspondence to:** Marta Cabrera. Department of Philosophy, Universitat de València. Avd. Blasco Ibáñez, 30 (46010 Valencia-Spain) – Marta.Cabrera@uv.es – <https://orcid.org/0000-0001-6702-3707>

How to cite: Cabrera, Marta (2022). «Duchenne smiles are actions not mere happenings: lessons from the debate on expressive action»; *Theoria. An International Journal for Theory, History and Foundations of Science*, 37(2), 163-179. (<https://doi.org/10.1387/theoria.22572>).

Received: 2021-03-05; Final version: 2022-04-25.

ISSN 0495-4548 - eISSN 2171-679X / © 2022 UPV/EHU



This work is licensed under a
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

In this paper, I will argue that, contrary to what is generally assumed in the debate on expressive action, we do not have good reasons to exclude facial and bodily expressions of emotion such as smiling, frowning, crying, clenching one's fists or hanging one's head from the category of actions.¹ For this purpose, I will compare facial and bodily expressions of emotion with simple expressive actions, such as jumping for joy, covering one's face in shame, scratching one's head in frustration or punching the air in delight. I will try to show that simple expressive actions cannot be presented as actions while excluding facial and bodily expressions of emotion from this condition. My contention will then be that either both sorts of behaviour are to be identified as actions or neither is. The latter sounds rather implausible, though, as we would have to assimilate jumping for joy or covering one's face in shame to spasms, which conflicts with the way we relate to such behaviours. My conclusion will then be that both simple expressive actions and facial and bodily expressions of emotion should be included within the category of actions, at least on the basis of the main assumptions in the current debate on expressive action.

In section 1, I introduce the category of facial and bodily expressions of emotion, together with the reasons why participants in the debate have tended to exclude such behaviours from the category of actions. In section 2, I present the category of simple expressive actions and briefly examine the way such behaviours have been accounted for in the literature. In section 3, I compare simple expressive actions with facial and bodily expressions of emotion and argue that, on the basis of the main assumptions in this debate, either both kinds of expressive behaviour are to be identified as actions or neither is. Since the latter sounds rather implausible —given the way we approach behaviours such as jumping for joy or covering our faces in shame— I conclude that both simple expressive actions and facial and bodily expressions of emotion should be included within the category of actions.

1. Facial and bodily expressions of emotion

Within the debate on expressive action, there is a wide consensus with respect to the idea that *facial and bodily expressions of emotion* such as smiling or frowning are not actions of an agent (Hursthouse, 1991; Goldie, 2000; Betzler, 2007; Helm, 2016; Bennett, 2016; Müller & Wong, forthcoming).² The introduction of such category stems from the need to clarify which kinds of expressive behaviour are actions and which are mere happenings an agent

¹ By “debate on expressive action” I refer to the debate initiated by Hursthouse with her paper *Arational actions* (1991). This debate has been variously referred to in the literature as the debate on arational actions, emotional behaviour, actions by emotion, emotional expression or expressive action (Goldie, 2000; Döring, 2003; Betzler, 2007, 2009; Helm, 2016; Bennett, 2016, 2021; Müller & Wong, forthcoming). Following Betzler and Bennett, I will talk about the debate on expressive action when I speak of such debate.

² In Hursthouse’s terms, facial and bodily expressions of emotion are not *intentional* actions. Throughout this debate, participants have indistinctively talked about “actions” and “intentional actions”, and have distinguished them from what merely happens to an agent (Goldie, 2000; Betzler, 2007; Müller & Wong, forthcoming), mere evincings (Helm, 2001, 2016), reflexes (Bennett, 2016) or reactions (Bennett, 2021). In this paper, rather than the contrast between intentional and non-intentional action, I will follow the contrast between *actions and happenings* that appears to underly all such proposals.

undergoes. As it is evident, whether or not some piece of behaviour is an expression of our agency is not a trivial issue: it is one of the most important distinctions in philosophy and everyday life, for it is clear that we relate very differently to actions than to mere happenings.

The clearest statement of the idea that behaviours such as smiling or frowning cannot be seen as actions may be found in Goldie's description of the category of facial and bodily expressions of emotion:³

I turn now to those expressions of emotion *which are not in any sense actions*. I have in mind not only facial expressions of emotion (the smile and the frown, the contortion of the face in fear, the opening wide of the eyes in surprise), but also laughter, the flow of tears and the tremor of fear in the voice. They too, like bodily changes, can be *causally explained*, but they are not actions, in spite of the fact that we think of them as expressions of emotion, rather than as part of the emotion itself. In this category I am mainly interested in those involuntary bodily movements which are involved in expressing some emotions. I emphasize *involuntary* in order to contrast them with what we can, and sometimes do, directly try to do. We can smile in order to give the impression we are glad when we are not; we can smile when we are glad in order to show that we are; and we can smile because smiling gives us pleasure. But all these ways of smiling are to be contrasted with the *genuine* or "Duchenne" smile, which involves distinct muscles which we cannot directly try to move (...) we need not feel pushed towards postulation of a motive for his smiling, or of a *means-end belief-desire explanation* of it. A smile would be an action, explicable like this, only if it were of the sort discussed at the end of the preceding paragraph [when we smile in order to give the impression we are glad when we are not]. (Goldie, 2000, p. 34-5, emphasis added)

The idea here is that facial and bodily expressions of emotion are not always performed in view of some particular end – they often simply express or reveal the emotion that someone is feeling, in which case, and for the reasons emphasised above, they cannot be regarded as actions of an agent. When speaking of this kind of behaviour then, we have in mind *genuine* facial and bodily expressions, and not crocodile tears or the smiles that we perform "in order to give the impression we are glad when we are not" (Goldie, 2000, p. 34)

Goldie adds that we should distinguish such kind of expressive behaviour from what he calls "bodily changes", that is, those alterations that take place when we experience emotions and that can never be regarded as actions: "autonomic nervous system responses and hormonal changes such as sweating, change of heart-rate, secretion of adrenalin and so forth, and muscular reactions such as trembling, flinching, (...) These changes just happen to us; they are not things which we do or can directly try to do" (2000, p. 26). According to him, then, both bodily changes and facial and bodily expressions of emotion are *involuntary reactions* (and not responses) *that have to be causally explained*. The only difference between these two categories seems to be that facial and bodily expressions of emotion are *expressions* of the emotion while bodily changes are *part of the emotion itself*. Goldie does not explain, however, how we are to understand such a contrast.⁴

³ Although Goldie does not give a name to this category, "facial and bodily expressions of emotion" seems to be a convenient label given the kind of behaviour included in such category.

⁴ The distinction between facial and bodily expressions of emotion, on the one hand, and bodily changes, on the other, together with the idea that none of these kinds of behaviour are actions is fairly widespread within the debate on expressive action. Müller and Wong, for instance, distinguish between "basic bodily expressions" —smiling, frowning or dropping one's voice— and "reflex-like

Since our focus in this paper is on facial and bodily expressions of emotion, we will leave bodily changes out of our discussion. We may summarise the reasons for the exclusion of facial and bodily expressions of emotion from the category of actions by reference to Goldie's first quote (2000, p. 34-5):

- C1. They cannot be explained by a *means-end belief-desire explanation* or, in other words, they are not instrumental.
- C2. They are *causally explained*, that is, they are alien to rationality.
- C3. They are *involuntary*.

In light of these criteria, we may conclude that, from this view, smiling, frowning, crying, laughing, opening one's eyes, hanging one's head, clenching or pounding one's fists, etc. are analogous, in a relevant sense, to *spasms and reflexes*: they are mere happenings or symptoms disconnected from our agency.⁵ That is, they are not *things we do*, but rather *things that happen to us*.

This view, however, seems to raise an important dilemma within the debate on expressive action. As I will try to show, the claim that facial and bodily expressions of emotion are *not actions* —following C1-C3— conflicts with the claim that simple expressive actions, such as jumping for joy or covering one's face in shame, are *actions*.⁶ As I will argue, C1-

emotional reactions” —recoiling in shock or disgustedly spitting out an insect—and claim that none of these kinds of behaviour are actions because they are *not instrumental nor rational nor voluntary* (forthcoming). Similarly, and following Hursthouse, Bennett distinguishes between “phenomena like blushing, sweating” and “forms of behaviour like smiling”. He claims that none of these kinds of behaviour are actions and that the difference between them is that the latter “often occur involuntarily but (...) it is possible to stop or to suppress [them]” (2016, p. 75), unlike in the case of blushing or sweating. This is precisely Hursthouse’s idea: she says that we often clench our fists, smile or frown without being aware of such movements—in which case they are not actions of ours— while, on some other occasions, we become aware of them while we are already performing them (1991, p. 65). In such case, she says, the thing *we do* is simply *refraining from stopping*. From this view, then, genuine smiles can never be regarded as actions—they are mere “manifestations of emotion [that] remain immune to rational explanation” (Bennett, 2016, p. 75). This is why Bennett claims that we should “distinguish the *symptoms* of some mental state, such as crying or smiling, from those objects (and, I claim, actions) that possess expressive power because they seem to capture or reflect some mental state (or the content thereof)” (2016, p. 84). By contrast, Helm seems to conflate facial and bodily expressions of emotion and bodily changes in a single category, which he labels “evincings”. Within such category, we find “bits of behaviour that are simply caused by the emotion, behaviours such as changes in respiration or pulse rate or certain facial or other gestures, such as a wince or fist pounding” (2016, p. 97) as well as behaviours such as gritting one’s teeth or hanging one’s head in shame (Helm, 2001, p. 75). This kind of behaviour should be contrasted with what Helm calls “expressions of emotion”, that is, “actions that are rationally motivated by the evaluative content of the emotion”, such as jumping for joy or rumpling someone’s hair out of love (2016, p. 97).

⁵ My claim is not that participants in this debate hold that facial and bodily expressions of emotion are *identical* to spasms and reflexes, but, rather, that they grant them *the same status as mere happenings*.

⁶ Someone might wonder whether C1-C3 are necessary or sufficient conditions for a behaviour to count as an action, or whether they are separate conditions at all. I myself have no clear answer to this question, however, whether C1-C2 are necessary or sufficient conditions for agency does not affect my argument, since my claim is that bodily and facial expressions of emotion and simple expressive actions *satisfy the same criteria* and, therefore, should be included within the same category, regardless of whether such criteria impose necessary or sufficient conditions.

C3 cannot justify our classifying these two kinds of expressive behaviour differently, for it seems that:

- Neither can be instrumentally explained.
- Both can be rationally explained.
- Both can involve muscles we cannot directly control.

If the arguments that I will present in favour of such view are sound, it seems that participants in the debate on expressive action are faced with an important dilemma: *either both sorts of behaviour are to be identified as actions or neither is*.

As I will argue in coming sections, the latter sounds rather problematic, as we would have to assimilate simple expressive actions, such as jumping for joy or covering one's face in shame, to spasms, which conflicts with the way we relate to such behaviours. I will conclude, then, that —on the basis of the main assumptions in the current debate— both simple expressive actions and facial and bodily expressions of emotion *should be included within the category of actions*. Insistence on the idea that facial and bodily expressions of emotion cannot be regarded as actions, would seem to require that new reasons should be provided to help differentiate such pieces of behaviour from simple expressive actions.

However, before we can compare facial and bodily expressions of emotion with simple expressive actions and argue for the idea that they should both be included within the category of actions, we must examine in some detail what simple expressive actions are and the way they have been accounted for in the debate on expressive action.

2. Simple expressive actions

At first sight, jumping for joy seems like one of the simplest things we can do. Compared to actions such as cooking a meal, writing a novel or building a bridge, happily jumping up and down when receiving good news seems to be an action that does not require from us a too complex analysis. The same goes for something apparently as simple as covering one's face in shame —when remembering, for instance, a particularly shameful moment— or scratching one's head in frustration.⁷

Such behaviours, however, as Hursthouse (1991) famously noted, cannot be regarded as actions by the lights of the standard account of agency. According to such account, (1) a behaviour counts as an action if it can be rationalised, and (2) to rationalise an action is to present it as something instrumentally valuable from the perspective of attaining some end. That is, if A is an action performed by subject S, we should be able to explain A by offering a desire-belief pair (D-B) such that satisfying desire D is the aim of S and S has the belief B according to which A allows her to satisfy D.⁸ Running away from a predator in fear, for instance, clearly admits of a rationalisation of this kind and, consequently, can be considered an action: the subject has the desire to protect herself and the belief that running away

⁷ We are considering cases in which an agent *genuinely* jumps for joy, covers her face in shame or scratches her head in frustration. In such cases, the agent does not perform such behaviours in view of some end – she does not jump, for instance, in order to communicate her happiness; she just joyfully jumps at some good news she has received, for instance.

⁸ See Davidson (1963).

is the best means of satisfying such desire. If we take this view, therefore, we cannot regard jumping for joy or covering our faces in shame as actions: when we jump for joy, we are not trying to instrumentally satisfy an end through our jumping. Likewise, when we cover our faces in shame, we are not trying to instrumentally satisfy an end through our covering our faces. According to this, and against our initial intuition, then, it does not look like there is *something we do* in such cases – such behaviours are mere happenings we undergo.

The standard account of agency, though, was put into question by Hursthouse, who defended the idea that there is a class of actions —*arational actions*— which are *clearly* actions despite not being instrumentally rationalisable.⁹ This would be the case of jumping for joy, covering one's face in shame, rumpling the hair of a loved one out of affection, rolling in grief in the clothes of a loved one who has died, destroying things in a fit of rage, gouging out the eyes of a rival in a picture out of hatred, etc. In all these cases, we are unable to explain what the agent does by reference to a belief-desire pair that shows her action as something instrumentally valuable. According to Hursthouse, these actions can only be explained by saying that “in the grip of the emotion, the agent just felt like doing them” (1991, p. 61). From this view then, we can admit that some of the things we do are disconnected from instrumental rationality —and are, hence, arational— but are still actions of ours. If this is correct, it seems that the standard account of agency is in trouble and that our intuition that jumping for joy and covering one's face in shame are *actions* can be preserved.

Many have tried, nonetheless, to show that the standard account is immune to Hursthouse's attack because the class of actions that she presents —or, at least, most of the actions she presents— *can* be rationalised after all (Smith, 1998; Goldie, 2000; Helm, 2001, 2016; Betzler, 2007, 2009; Döring, 2003; Scarantino & Nielsen, 2015; Bennett, 2016, 2021; Müller & Wong, forthcoming). The shared thought here seems to be that, if some piece of behaviour can be regarded as an action, then such behaviour *must be* connected to rationality in some sense. From this point of view, then, the idea that there is something like *arational actions* cannot be correct.

Before examining the way in which some such views have accounted for the behaviours we are interested in —jumping for joy, covering our faces in shame, etc.— we should briefly sketch two crucial features of how the debate has evolved since Hursthouse's seminal paper. Firstly, the current debate has produced a number of distinctions within the class of so-called arational actions. Although not everyone explicitly agrees in this respect, the common assumption is that not all cases in Hursthouse's list belong to the same kind (Goldie, 2000; Betzler, 2007, 2009; Scarantino & Nielsen, 2015; Pineda, 2019; Müller & Wong, forthcoming) and should be divided into three types: symbolically displaced actions, radically displaced actions and simple expressive actions.¹⁰ The cases we are inter-

⁹ In Hursthouse terms, such actions are clearly *intentional* actions.

¹⁰ Symbolically and radically displaced actions are actions in which the object of an emotion is displaced to a different object that emerges as the explicit object of the action. In the case of *symbolically displaced actions*, “there is some relation between object of emotion and object of expressive action onto which the agent's mind might latch” (Goldie, 2000, p. 31). This is what happens when one rolls in grief in the clothes of a loved one who has died or when gouging out the eyes of a rival in a picture out of hatred. Since the object of the agent's emotion is not available for her to act on it —in the first case, because the loved one has passed away and, in the second, because there are social conventions discour-

ested in —jumping for joy, covering one's face in shame, punching the air in delight or scratching one's head in frustration— belong to the latter type. Secondly, attempts to rebut Hursthouse's case against the standard account have mainly focused on symbolically displaced actions.¹¹ As a result of this, simple expressive actions and radically displaced actions have gone under-explored. In fact, simple expressive actions tend to be presented as *residual, problematic or even borderline cases* (Goldie, 2000; Betzler, 2007, 2009; Bennett, 2016). The usual strategy has been to show that symbolically displaced actions *do have a purpose* and, therefore, that we can easily rationalise them in instrumental terms.¹² Hence, such kind of behaviours are not *arational* actions – they are actions perfectly explained by the standard account. Furthermore, some of the participants in the debate seem to be convinced that this line of reasoning not only applies to symbolically displaced actions, but to all actions included in Hursthouse's list and, therefore, that her case against the standard account of agency fails.

This last strategy, however, seems to have devastating consequences for simple expressive actions, since instrumental explanations of behaviours such as jumping for joy or covering one's face in shame clearly *over-intellectualise* them. Bennett argues, for instance, that so-called arational actions do have “*a purpose* —that of doing justice, or giving adequate external form to one's sense of the situation” (2016, p. 74, emphasis added)— which means that they can be instrumentally rationalised – and that even the “spontaneous case [jumping for joy] is susceptible of a ‘reading’ that might show why it would be appropriate to select [such expressive behaviour] as a fitting vehicle for that emotion if (...) one is in the business of deliberatively selecting a vehicle for one's emotion” (2016, p. 91-2). The problem is that it does not look like we deliberatively select a vehicle for our joy when we jump for joy.

Smith (1998), on the other hand, does not mention examples of simple expressive actions but assumes that his proposal can explain away all cases presented by Hursthouse.

aging us from gouging people's eyes out— her action takes a different object that bears some symbolic relation with the object of her emotion. There is also a displacement of the object of one's emotion in *radically displaced actions* but such displacement involves no symbolic identification. The connection between the object of one's emotion and the object of the expressive action is arbitrary, as when one destroys an object in a fit of rage or kicks a chair out of anger. In such cases, it seems that any other object at hand —a table, a bag, etc.— could have just as well been the object of one's action. Finally, *simple expressive actions* are those actions where neither the object of one's emotion is displaced nor a symbolic identification is at issue. These are cases such as jumping for joy or hiding one's face out of fear or in shame. I have borrowed the terminology “symbolically displaced action” and “radically displaced action” from Scarantino and Nielsen (2015) and “simple expressive action” from Pineda (2019).

¹¹ Gouging out the eyes of a rival in a picture out of hatred (Hursthouse, 1991; Goldie, 2000; Döring, 2003), rolling in grief in the clothes of a loved one who has died (Smith, 1998) and carrying one's grandfather's coffin during his funeral out of love (Bennett, 2016), for instance. An exception to the general tendency can be found in Helm (2001, 2016) and Müller and Wong (forthcoming).

¹² Not everyone agrees that the mental states that rationalise the agent's behaviour are beliefs and desires. Döring (2003) and Scarantino and Nielsen (2015), for instance, argue that the end of the agent's action is provided by the emotion she experiences and not by a desire that she may have. It will be enough for our discussion, though, that all of them defend that symbolically displaced actions have *a means-end structure* in virtue of which such actions can be rationalised and adequately accounted for as *actions* by the standard account. See Bennett (2021) for an exception to this tendency.

According to him, when a man rolls in grief in the clothes of his dead wife “the man is doing what he is doing because he desires to roll around in his dead wife’s clothes and believes that he can do so by doing just what he is doing; that is, by rolling around in those particular clothes that he is rolling around in” (1998, p. 22). Smith adds that in order to understand the desire of the agent, we must make reference to his emotion, but that the belief-desire pair suffices to rationalise his action. If we turn to the case of jumping for joy, however, it does not seem that one jumps for joy when receiving good news because one desires to jump and believes that one can do so by just doing what one is doing, that is, jumping. Such an explanation does not look like a good description of what goes on in this case.¹³

Given these difficulties, Goldie (2000) and Betzler (2007, 2009) have speculated that, contrary to our intuitions, simple expressive actions might not be actions at all. Goldie says that “jumping for joy, scratching your head in frustration, and punching the air in delight (...) [are] things which one does [that] are unlike the genuine smile in that the latter involves a movement of certain muscles which one cannot directly try to move. But still, surely a genuine spontaneous jump for joy [...] no more involves a belief than does the genuine spontaneous smile” (2000, p. 36). In a similar line, and after concluding her alternative account of so-called arational actions, Betzler wonders whether her ideas could be extended to simpler kinds of behaviour:

In contrast, hiding one’s face from fear, blushing in shame, jumping for joy, or posturing in front of adversaries are cases that might be thought to be intelligible without assuming any particular point of view of the respective agent. Some of these expressive actions may be in the vicinity of *mere bodily movements*. Others can be explained simply by referring to the instinct of survival that all agents share as members of a species. A concession in such cases, however, is compatible with the rationalization applying in many other cases. (Betzler, 2009, p. 284-5, emphasis added)

These ideas, as I noted, are merely speculative and stem from the view that what we have been calling simple expressive actions are residual and problematic cases. Assimilating such behaviours to mere happenings, though, seems to have some rather implausible consequences that should be properly addressed by anyone who wants to defend this view. In particular, if jumping for joy or covering one’s face in shame were nothing more than mere happenings —or, in Betzler’s terms, mere bodily movements— we would need an explanation of the way in which we usually approach such behaviours. When we see that someone who has just received some news starts jumping up and down, we immediately understand that the news she has received is *good* – or, at least, that it is good *from the agent’s point of view*. This means that we approach her behaviour as a response to something that she evaluates as worth celebrating. Her response, moreover, can be assessed as appropriate or inappropriate, or as proportional or disproportional. In fact, if jumping were, for whatever reasons, inappropriate in the given circumstances, the subject could be *blamed* for not having waited for a more convenient time to express her happiness. As we can see, though, this is not the way we relate to mere happenings —which are disconnected from our agency—

¹³ Smith’s proposal has, in fact, been widely rejected as a solution to Hursthouse’s problem since it does not seem to provide a good rationalisation of any of the cases presented in her list (Goldie, 2000; Döring, 2003; Betzler, 2007; Bennett, 2016; Pineda, 2019; Müller & Wong, forthcoming).

and therefore Goldie's and Betzler's suggestions do not seem to provide, in principle, a good explanation of these cases.

To recapitulate, we started by saying that —compared to actions such as cooking a meal, writing a novel or building a bridge— jumping for joy, covering one's face in shame, punching the air in delight or scratching one's head in frustration seem to be among the simplest actions we can perform. This intuition, however, seems to be questioned by the standard account of agency, according to which, behaviour that cannot be instrumentally rationalised cannot be regarded as the actions of an agent. Hursthouse's critique of the standard account, however, allows us to regard such cases as actions, in particular as actions that are disconnected from rationality and that can only be explained by saying that "in the grip of the emotion, the agent just felt like doing them" (1991, p. 61).

This last idea is nonetheless rejected by many participants in the debate who try to show that the cases presented by Hursthouse can be rationalised and thus be considered actions by the lights of the standard account. As we have seen, however, instrumental explanations of simple expressive actions tend to over-intellectualise them, therefore explanations of this kind do not succeed in showing that jumping for joy or covering one's face in shame are not *arational* actions. Alternatively, a defendant of instrumental rationality can question whether such behaviours are actions at all —since they do not involve any means-end belief-desire pair— but this leads us to a very difficult position: jumping for joy or covering one's face in shame are assimilated to spasms and reflexes, which conflicts with the way we deal with such behaviours.

It appears, then, that simple expressive actions are still problematic cases for the standard account: we have the intuition that they are actions but they seem to be disconnected from rationality – or, at least, *from the kind of rationality presupposed in the discussion so far*. As Helm (2001, 2016) and Müller and Wong (forthcoming) have convincingly argued, all previous views assume that rational action is necessarily *instrumentally* rational action. However, they claim, an alternative conception of rationality is needed if we are to explain the sense in which some expressive behaviours are not a matter of mere accident:¹⁴

We should not think that all intentional action is instrumental, done for the sake of achieving some end, for *all that's needed for an action to be rationally explicable is that it has a point the explanation reveals to be worthwhile* (...) the point of jumping for joy is that it is a celebration, and in feeling joy one feels its target to be a good worthy of celebration (...) kissing someone or rumpling his hair or, depending on the circumstances, holding him or even just sitting quietly with him all have a point in that they are all ways of being solicitous, and in feeling occurrent love one feels someone as worthy of solicitude. And tearing one's hair or clothes has a point insofar as these are all ways of mourning, and to feel grief is to feel a loss to be worthy of mourning. In each of these cases, the emotional expressions are *not means* undertaken to achieve the end of celebration, solicitude or mourning; *they are the celebration, solicitude and mourning*. (Helm, 2016, p. 98, emphasis added)

The rejection of instrumentality as a criterion for agency, then, allows us to see a way in which simple expressive actions might be connected to rationality – and are, therefore, *not*

¹⁴ See also Bennett (2021) for a non-instrumental explanation of expressive action. In said paper, however, he exclusively focuses on cases of symbolic action, such as Achilles' dragging Hector's body around the walls of Troy or the kneeling of the believer in church.

arational actions: “[e]xplanations of this kind articulate a further aspect of our ordinary conception of action: in contrast to mere behaviour, an action is rational in this reason-responsive sense” (Müller & Wong, forthcoming). From this view, Hursthouse’s account of behaviours such as jumping for joy or covering our faces in shame should be rejected on the basis of the following assumption:

Non-instrumental Rationality Assumption: Jumping for joy —as well as the rest of our simple expressive actions— can be regarded as an action because it has a point that shows that the agent is responding to reasons when she performs such movements.

This seems to be the reason why it is so hard to view simple expressive actions as analogous to spasms and reflexes. In the latter cases, the agent is clearly not responding to reasons — spasms and reflexes do not have a point—, that is, there is nothing that, from the agent’s point of view, counts in favour of her movements. Although this might not be the only kind of non-instrumental rationality that could explain the sense in which behaviours such as jumping for joy or covering one’s face in shame are actions, the proposals of Helm and Müller and Wong seem to have correctly identified that instrumentality is an obstacle for our understanding of an important part of our expressive behaviour.

Once we have made room in the debate for simple expressive actions, I will argue that, on the basis of C1-C3, simple expressive actions and facial and bodily expressions of emotion cannot be classified differently, that is, with the former as actions and the latter as mere happenings. As we have seen, participants in this debate differ with respect to their understanding of simple expressive actions, however, *both critics and supporters of the instrumental model of rationality seem to agree that facial and bodily expressions of emotion are different from simple expressive actions and cannot be regarded as actions*.¹⁵ In what remains of the paper, we will examine the reasons provided in favour of this view.

3. Facial and bodily expressions of emotion versus simple expressive actions

As we saw in section 1, participants in the debate on expressive action tend to exclude facial and bodily expressions of emotion from the category of actions on the basis of the following criteria:

- C1. They are not instrumental.
- C2. They are alien to rationality.
- C3. They are involuntary.

I will now provide three arguments (A1-A3) in favour of the idea that, according to C1-C3, either *both* simple expressive actions and facial and bodily expressions of emotion *are to be identified as actions or neither is*. Since the latter sounds rather implausible, given the way we approach simple expressive actions, I will conclude that both simple expressive actions

¹⁵ As we have seen, however, Goldie and Betzler raise the possibility that simple expressive actions might be like facial and bodily expressions of emotion qua happenings, but, as we have already argued, this consequence is hard to accept.

and facial and bodily expressions of emotion should be included within the category of actions. In order to simplify the following discussion, I will mainly compare the case of smiling with the case of jumping for joy.

- (A1) *If a piece of behaviour must be instrumentally rational in order to count as an action, then facial and bodily expressions of emotion are not actions but neither are they simple expressive actions.*

According to the first criterion, facial and bodily expressions of emotion cannot be regarded as actions because there are no belief-desire combinations that could instrumentally explain why the agent performs them. As we saw, the distinctive feature of *genuine* smiles is, precisely, that the subject does not smile in view of some end. However, if we accept that smiles are not actions because in order to count as an action a piece of behaviour must be instrumentally rational, then simple expressive actions must also be excluded from the category of actions: as it is assumed in the debate, it does not look like we are trying to instrumentally satisfy an end through our jumping when we jump for joy.

According to the *Non-instrumental Rationality Assumption*, though, we should resist the idea that jumping for joy or covering our faces in shame are analogous to spasms and reflexes qua mere happenings we undergo, *even though we are unable to explain them in instrumental terms*. As Helm's and Müller and Wong's non-instrumental conception of rationality shows, we seem to have alternative ways of approaching such behaviours as actions. If, following this view, we accept that jumping for joy is an action *regardless* of its lacking a means-end structure, then we should not exclude smiling from the category of actions just *because* it lacks a means-end structure.¹⁶

- (A2) *If an expressive action can be rational just because it has a point, without being instrumentally rational, as in the case of simple expressive actions, then facial and bodily expressions of emotion should not, in principle, be excluded from the category of actions*

As we saw before, according to the *Non-instrumental Rationality Assumption*, simple expressive actions “can have *a point* which makes them intelligible as reason-responsive, even if we cannot make sense of them as instrumental” (Müller & Wong, forthcoming) and “inasmuch as the performance of certain movements has a point for the agent, it seems not to be a matter of mere accident” (Müller & Wong, forthcoming). Such proposal allows us, then, to regard such behaviours as actions and to rationally explain them.

According to Müller and Wong and Helm, however, this is an explanation that *does not apply to facial and bodily expressions of emotion*: such behaviours are alien to rationality and, therefore, must be *causally explained* (see also Goldie, 2000). Regarding smiles and frowns, Müller and Wong hold that there are not “any cognized aspects of [the] situation

¹⁶ As it is evident, this conclusion is conditional on the correctness of the *Non-instrumental Rationality Assumption*. Although such assumption is not fully defended in this paper, it appears to be plausible given our previous discussion on simple expressive actions. As will be developed in the following, my main claim is that, on the basis of the *Non-instrumental Rationality Assumption*, reflection on the place that facial and bodily expressions of emotion have in our lives shows that they too should be regarded as actions. I thank an anonymous referee for encouraging me to stress this more clearly.

that intelligibly motivate them. Rather, they are simply triggered by those circumstances” (forthcoming). Similarly, Helm claims that

although you can tremble from fear or hang your head in shame at otherwise inopportune moments, there is nothing that would count as trembling or hanging one’s head *correctly or incorrectly*. Because such behaviors are without a point, there is nothing in the content of the emotion that makes them intelligible by saying something in favor of behaving in these ways [...] such behaviors are *disconnected from rationality*.¹⁷ (Helm, 2001, p. 75, emphasis added)

This seems to be Bennett’s view too, in particular, when he says that we should “distinguish the *symptoms* of some mental state, such as crying or smiling, from those objects (and, I claim, actions) that possess expressive power because they seem to capture or reflect some mental state (or the content thereof)” (2016, p. 84) and that only the latter raise “issues of adequacy, appropriateness and inappropriateness that ground normative assessment of different forms of expression” (2016, p. 84).

These proposals seem to be based on the idea that examples of behaviour that are mere happenings we undergo, such as spasms or reflexes, are not subject to rational standards. That is, such behaviours *cannot be evaluated as good or bad responses to the situation in which a subject finds herself*. This seems to be clearly correct in the case of spasms: we can see how the subject’s bodily movements have nothing to do with her evaluative perspective on the situation – consequently, her movements cannot be intelligibly evaluated as a good or bad response to her situation. Think of someone who is pathologically laughing all the time or someone who, due to a paralysis of some facial muscles, has a permanent smile on her face. Such movements just happen, indiscriminately – they are not responsive to anything and, thus, neither proportional or disproportional, appropriate or inappropriate. This is the reason why such behaviour can only be *causally* explained.

The question we may ask ourselves now is whether facial and bodily expressions of emotion *are analogous to spasms in this respect*, as suggested by the previous authors. For this purpose, let us introduce an example about smiles. Imagine a child who is playing with her toys, supervised perhaps by some family member, when her mother walks into the room after returning from work. Imagine the way the child turns to her mother with a beaming smile on her face and how the mother smiles with joy and love at the sight of her child. This smile is the kind of smile that the child, as she grows older, will continue to seek from her mother.

What can we say about the smiles in this example? According to the view we are examining, smiles are mere happenings we undergo, so the way we approach smiles —as well as the rest of our facial and bodily expressions of emotion—¹⁸ should not be very different

¹⁷ Such an idea contrasts with the fact that Helm regards crying out of sadness as an expressive action and describes its point as ‘mourning’ (2001, p. 77).

¹⁸ Although we are using an example about smiles in order to illustrate the idea that we relate to facial and bodily expressions of emotion differently than to spasms, we could easily examine similar examples in which the emotion and the facial and bodily expressions are different. “My eyes moved from child to mother as the girl shifted from barely audible whispers to choked admissions to hoarse gasping sobs. I noted that the mother’s face functioned as a vague mirror of her child’s. When Alice spoke softly, Ellen leaned forward, her eyes intent as her lips registered every insult with tiny movements. When Alice

from the way we approach spasms.¹⁹ This, however, does not seem to be the case, for (a) we relate to facial and bodily expressions of emotion as *responses* to something. This is the reason why we ask questions like “what makes you so happy?” or “why are you smiling?”. These questions presuppose an internal connection —a connection of intelligibility— between a smile and a situation, and not a connection of causal trigger. That is, we expect answers to these questions that mention aspects of a situation that merit a certain response or recognition by part of the subject and not just physiological or adaptive considerations. Indeed, it seems that we cannot merely view the smiles of the mother and the child as brute effects of their circumstances: such behaviour *makes sense* or is intelligible *given their situation* – it makes sense in a way in which banging one’s fist three times against the wall at the sight of one’s daughter, for instance, does not.²⁰

Moreover, and connected to the previous point, (b) we regard facial and bodily expressions of emotion as *revealing of the subject’s evaluative perspective* on the situation.²¹ When we see the child’s smile, it is clear for us that she apprehends her mother’s smiling arrival as a happy event (the happiest of the day, perhaps) and vice versa, the sight of her child smiling at her is apprehended by the mother as a joyful and moving scene. The idea here is that, in someone’s smile, we seem to *see or anticipate the person*, that is, we see who she is and what she values. In this respect, smiles seem to differ from reflexes or spasms: the latter appear to be “disconnected” in an important sense from the person experiencing them, since, as we said, there seems to be no evaluation of the situation on the part of the subject in such cases.

Relatedly, there seem to be (c) *moral considerations* that regulate our smiles. When we say things like “your smile was quite insensitive” or “why don’t you smile? Don’t you see how important this is to her?” we seem to be blaming the subject in a different way in which we would blame someone for having a spasm or a reflex. The most we could say to someone who has undergone a spasm in a very inappropriate moment —imagine that this person suffers from a nervous disorder— is that she should not have put herself in a situation in which this was likely to happen. Smiles, however, seem to be subject to a different kind of considerations: if the mother in our example had not smiled at the smile of her child —if, stressed and exhausted from work, she had turned away from her— it seems like the child would have had a claim on her, that is, a claim to be seen and acknowledged by her as someone worthy of love and attention. It is from this last point of view, moreover, that we can understand the sense in which the child may continue to seek her mother’s smile through the years. Furthermore, (d) smiles seem to be open to *education*: we are educated at what to smile and how we smile.

cried, Ellen’s eyes grew smaller, a wrinkle appeared between her brows, and her mouth tensed into a thin straight line, but she did not weep. Maternal listening is of a special kind” (Hustvedt, 2011).

¹⁹ If the view we are examining were correct, a close look at our practices should reveal that we deal with or respond to spasms and smiles analogously. See Strawson (1962).

²⁰ In order to make sense of such strange behaviour we would need a *story* that explains the sense in which banging one’s fist three times against the wall is a way of acknowledging and rejoicing at the presence of one’s daughter (Berlin, 1958; Corbí, 2012).

²¹ I have borrowed the expression “revealing of the subject’s evaluative perspective on the situation” from Bennett (2021, p. 11), although he is referring to a much more complex kind of expressive behaviour (Achilles’ dragging Hector’s body around the walls of Troy).

This way of approaching smiles, however, seems to be at odds with the idea that they are mere happenings that are disconnected from rationality.²² The four interconnected features that we have just described suggest that facial and bodily expressions of emotion clearly raise issues of adequacy, appropriateness and inappropriateness: if there were not correct and incorrect ways —or proportional and disproportional ways— of smiling, frowning, crying, hanging one's head, pounding one's fist, etc., it is hard to see how we could make sense of our usual way of approaching such behaviour. If the view defended by the aforementioned authors were correct, though, it would seem that when we say things like "this is such good news, why are you not smiling?" or "your smile was quite insensitive", we are mistaken in approaching our smiles as behaviours that can be corrected – or, at least, we are mistaken in thinking that we are regulating our smiles through this way of speaking. Assuming that we are mistaken in this respect, however, would involve a radical revision of our experience; to put it in Williams' terms, it "would involve a much vaster reconstruction of our sentiments and our view of ourselves than may be supposed" (1981, p. 22).²³

The idea that facial and bodily expressions of emotion are responses to a situation that a subject apprehends from a certain evaluative perspective and that can be assessed as appropriate or inappropriate, proportional or disproportional, casts doubt on the claim that such behaviours are disconnected from rationality. In fact, without committing ourselves to the proposals of Helm and Müller and Wong, we can ask ourselves whether the explanation they provide of simple expressive actions could also be applied to facial and bodily expressions of emotion. And, intuitively, it seems that we can say that such behaviours also *have a point*: smiling out of joy can be seen as *celebrating* or *rejoicing* (which is rational because joy apprehends its object as worth celebrating), frowning in disappointment can be seen as *disapproving* (which is rational because disappointment apprehends its object as worthy of disapproval), hanging one's head in dejection can be seen as *surrendering* (which is rational because dejection apprehends its object as worth surrendering to), crying out of sadness can be seen as *mourning* (which is rational because sadness apprehends its object as worthy of grieving), laughing out of amusement can be seen as *rejoicing* or *enjoying* (which is rational because amusement apprehends its object as worthy of amusement), etc.

If this is plausible, then, it does not seem like there is a relevant difference between jumping for joy and smiling with respect to rationality: in both cases, we seem to be dealing with non-instrumental behaviours that are, nonetheless, rational responses to a situation given the way in which the emotion of the subject apprehends such a situation. Therefore, it does not seem that facial and bodily expressions of emotion can be excluded from the category of actions just because they are disconnected from rationality. If simple expressive ac-

²² This seems to be, in fact, a view that Scruton (1986) has defended within a different debate: "Smiling must be understood as a response to another person, to a thought or perception of his presence, and it has its own intentionality. To smile is to smile at something or someone, and hence when we see someone smiling in the street we think of him as 'smiling to himself', meaning that there is some hidden object of his present thought and feeling. The smile of love is a kind of intimate recognition and acceptance of the other's presence - an involuntary acknowledgement that his existence gives you pleasure" (1986, p. 64).

²³ This is a remark that Williams offers regarding the debate on moral luck (1981).

tions are regarded as actions by virtue of their being rational responses, then so should facial and bodily expressions of emotion.

- (A3) *If actions must be voluntary, then facial and bodily expressions of emotion are not actions, but then neither are many examples of instrumental behaviour that, according to the standard account, should be classified as actions*

As we saw in previous sections, the claim that simple expressive actions are actions and facial and bodily expressions of emotion are mere happenings also appears to be grounded in a difference in the control that an agent may have over her bodily movements. The assumption underlying this claim is that, when performing an action, we have control over the muscles involved, that is, we are capable of moving our bodies in the way we do and to do so directly. Goldie's example of the Duchenne smile —which, according to him, is not in any sense an action— is aimed at showing precisely this (2000, p. 35). When our smile is genuine, the muscles we employ are different from those we use when we move our faces on purpose to form a smile. So, the divide between an action and a mere happening is thus defined by our capacity to have a direct control over the musculature in virtue of which a certain bodily change takes place.

However, consider now the action of running away in fear, full of adrenaline. Imagine that you are walking through a forest and that a bear suddenly appears. You are scared, want to protect yourself and believe that running through the space between two trees further to your right is the best way to escape, since the bear will not fit between them. So, this is what you do and your action is properly rationalised by the desires and beliefs described in the previous sentence. Now, as you perform this action, you run faster and jump higher than you are able to in normal circumstances, since you are full of adrenaline. You might even use muscles that you are unable to employ when, for instance, going for a jog. In fact, if asked later to replicate your running away at will, your body would not move in the same way – you would not be in a state in which, due to the adrenaline in your system, you are able to move your body in the way you did when running away in fear. This fact, however, does not prevent us from considering that your running away in fear was an action of yours. Hence, it seems that the inability to directly control all the muscles involved in our movements does not suffice to discard that facial and bodily expressions of emotion could count as actions.

According to A1-A3, then, C1-C3 do not justify classifying facial and bodily expressions of emotion and simple expressive actions differently, for:

- Neither can be instrumentally explained.
- Both can be rationally explained.
- Both can involve muscles we cannot directly control.

Consequently, either both sorts of behaviour are to be identified as actions or neither is. The latter, as we saw, seems rather implausible given the way we relate to behaviours such as jumping for joy or covering our faces in shame. Therefore, we may conclude that *both facial and bodily expressions of emotion and simple expressive actions should be included within the category of actions*. As previously stated, if someone wants to insist on the idea that facial and bodily expressions of emotion are not actions, she should provide new reasons that could help differentiate such behaviours from simple expressive actions.

4. Conclusions

I started by introducing the category of facial and bodily expressions of emotion, as well as the reasons that several philosophers have provided in favour of their exclusion from the category of actions. As we have seen, in the current debate on expressive action, behaviours such as smiling, frowning or crying are regarded as mere happenings we undergo because they are thought to be neither instrumental nor rational nor voluntary. I then presented the category of simple expressive actions and the various ways in which they have been accounted for in the literature. The main claim I have argued for —through A1-A3— is that simple expressive actions cannot be presented as actions while excluding facial and bodily expressions of emotion from this condition: neither of these two kinds of expressive behaviour can be instrumentally explained, both can be rationally explained and both can involve muscles we cannot directly control.

As a result, I have concluded that, on the basis of the main assumptions in the current debate on expressive actions, simple expressive actions and facial and bodily expressions of emotion cannot be classified differently: either both sorts of behaviour are to be identified as actions or neither is. Classifying behaviours such as jumping for joy or covering one's face in shame as happenings, however, conflicts with the way we relate to such behaviours. Therefore, I have concluded that both simple expressive actions and facial and bodily expressions of emotion should be included within the category of actions.

Acknowledgments

I am very grateful to Josep Corbí, Jordi Valor, Christopher Bennett, Jonathan Mitchell, Pia Campegiani, Carlos Moya and two anonymous referees for their valuable comments on earlier versions of this paper. I am also indebted to the audiences in the *Valencia Colloquium in Philosophy* and the *2nd VLC Philosophy Workshop: Sharing Reasons*.

This work has been supported by the research projects PID2019-106420GA-I00 and PID2020-119588GB-I00 and by the grant BES-2017-081537 funded by MCIN/AEI/10.13039/501100011033 and “ESF Investing in your future”.

REFERENCES

- Bennett, C. (2016). Expressive actions. In C. Abell & J. Smith (Eds.). *Emotional expression: philosophical, psychological and legal perspectives* (pp. 73-94). Cambridge: Cambridge University Press.
- Bennett, C. (2021). The problem of expressive action. *Philosophy*, 96(2), 277-300.
- Berlin, I. (1958). *The crooked timber of humanity*. Princeton: Princeton University Press.
- Betzler, M. (2007). Making sense of actions expressing emotions. *Dialectica*, 61(3), 447-466.
- Betzler, M. (2009). Expressive actions. *Inquiry*, 52(3), 272-292.
- Corbí, J. E. (2012). *Morality, self-knowledge and human suffering: an essay on the loss of confidence in the world*. New York: Routledge.
- Davidson, D. (1963). Actions, reasons and causes. *Journal of Philosophy*, 60(23), 685-700.
- Döring, S. (2003). Explaining action by emotion. *The Philosophical Quarterly*, 53(211), 214-230.
- Goldie, P. (2000). Explaining expressions of emotion. *Mind*, 109, 25-38.

- Helm, B. (2001). *Emotional reason: deliberation, motivation and the nature of value*. Cambridge: Cambridge University Press.
- Helm, B. (2016). Emotional expression, commitment and joint value. In C. Abell & J. Smith (Eds.). *Emotional expression: philosophical, psychological and legal perspectives* (pp. 95-114). Cambridge: Cambridge University Press.
- Hursthouse, R. (1991). Arational actions. *Journal of Philosophy*, 88, 57-68.
- Hustvedt, S. (2011). *The summer without men*. London: Picador.
- Müller, M. J. & Wong, H. Y. (forthcoming). Which emotional behaviors are actions?. In A. Scarantino (Ed.). *Routledge handbook of emotion theory*. London/New York: Routledge.
- Pineda, D. (2019). *Sobre las emociones*. Madrid: Cátedra.
- Scarantino, A. & Nielsen, M. (2015). Voodoo dolls and angry lions: how emotions explain arational actions. *Philosophical Studies*, 172, 2975-2998.
- Scruton, R. (1986). *Sexual desire: a philosophical investigation*. London: Weidenfeld and Nicholson.
- Smith, M. (1998). The possibility of philosophy of action. In J. Bransen (Ed.). *Human action, deliberation and causation* (pp. 17-41). Dordrecht: Kluwer.
- Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy*, 48, 187-211.
- Williams, B. (1981). *Moral luck*. Cambridge: Cambridge University Press.

MARTA CABRERA is a PhD candidate at the Universitat de València. Her main areas of interest are philosophy of emotions, philosophy of mind and action theory.

ADDRESS: Department of Philosophy, Universitat de València. Avd. Blasco Ibáñez, 30 (46010 Valencia-Spain). Email: Marta.Cabrera@uv.es. ORCID: 0000-0001-6702-3707

ARTÍCULO III: *¿Qué es una emoción?*

MARTA CABRERA
Universitat de València

¿Qué es una emoción?

What is an emotion?

Recibido: 22/7/2021. Aceptado: 7/10/2021

Resumen: El objetivo de este artículo es realizar un repaso de las teorías de las emociones más importantes de las últimas décadas y valorar en qué medida tales teorías ofrecen una buena caracterización del fenómeno en cuestión. Para alcanzar una comprensión adecuada de las diferentes teorías, así como del rumbo que ha tomado el debate actual sobre las emociones, es importante que prestemos especial atención a algunos de los compromisos filosóficos involucrados en la formulación de estas teorías. La revisión cuidadosa de algunos de estos compromisos nos llevará a plantearnos si algunas de las dificultades a las que se enfrentan varias de las teorías de las emociones actuales podrían tener origen en tales compromisos filosóficos y no, como suele pensarse, en la complejidad de las emociones.

Abstract: The aim of this paper is to review the most important theories of emotions in recent decades and to assess the extent to which these theories provide a good characterisation of the phenomenon in question. In order to achieve an appropriate understanding of the different theories, as well as of the direction that the current debate on emotions has taken, it is important that we pay special attention to some of the philosophical commitments involved in the formulation of these theories. Careful examination of some of these commitments will lead us to consider whether some of the difficulties faced by several of the current theories of emotions might stem from such philosophical commitments rather than, as is often thought, from the complexity of emotions.

Palabras clave: episodios, intencionalidad, fenomenología, propiedades evaluativas.

Keywords: episodes, intentionality, phenomenology, evaluative properties.

I. INTRODUCCIÓN

EL PRESENTE ARTÍCULO trata la cuestión de qué son las emociones desde la perspectiva de la filosofía de la mente actual. A pesar de que esta no sea una cuestión nueva en filosofía, pues durante siglos las emociones han constituido un importante tema de discusión filosófica (bajo diferentes denominaciones: pasiones, sentimientos, agitaciones, perturbaciones, afectos, etc.), esta es una cuestión que merece nuestra atención debido al creciente protagonismo que el estudio de las emociones ha adquirido en las últimas cuatro o cinco décadas. Este protagonismo se manifiesta de manera clara en el hecho de que, a día de hoy, distinguimos una subdisciplina específica llamada “filosofía de las emociones” dentro del ámbito de la filosofía de la mente cuyo objetivo principal consiste en esclarecer cuál es la naturaleza de las emociones. En otras palabras, la pregunta más básica a la que se enfrenta la filosofía de las emociones es “¿qué es una emoción?”. Esta es una pregunta que, en un sentido importante, precede a cualquier tipo de reflexión filosófica sobre las emociones, dado que la caracterización que ofrezcamos del fenómeno condicionará significativamente el tipo de cosas que podamos decir sobre las emociones.

En lugar de ofrecer una respuesta concreta a la pregunta “¿qué es una emoción?”, el objetivo de este artículo es realizar un breve repaso de las teorías de las emociones más importantes de las últimas décadas y valorar en qué medida tales teorías ofrecen una buena caracterización del fenómeno en cuestión. Para alcanzar una buena comprensión de las diferentes teorías, así como del rumbo que ha tomado el debate actual sobre las emociones, es importante que prestemos especial atención a algunos de los compromisos filosóficos involucrados en la formulación de estas teorías, y no sólo a sus argumentos a favor y en contra. La revisión cuidadosa de algunos de estos compromisos nos llevará a plantearnos si algunas de las dificultades a las que se enfrentan varias de las teorías de las emociones actuales podrían tener origen en tales compromisos filosóficos y no, como suele pensarse, en la complejidad de las emociones.

Comenzaremos aclarando en la segunda sección, por una parte, el modelo de explicación desde el que emprenderemos el estudio de las emociones y, por otra, el tipo de fenómenos afectivos que, a pesar de estar estrechamente conectados con las emociones, no son emociones y, por tanto, quedarán fuera de la presente discusión. A continuación, en la tercera sección, introduciremos los aspectos fundamentales de las emociones que toda teoría satisfactoria habría de ser capaz de explicar, a saber: su aspecto intencional, su aspecto evaluativo y su aspecto sentido o fenomenológico. En la cuarta sección, examinaremos las teorías de las emociones más importantes de las últimas décadas en el siguiente orden: teorías cognitivistas (teoría creencia-deseo y teoría juicialista), teorías

de la percepción y teorías del sentir (teoría jamesiana y teoría actitudinal). A lo largo de las distintas secciones, iremos identificando una serie de controvertidas cuestiones que condicionan de manera significativa el debate actual y determinan el tipo de respuesta que podemos dar a la pregunta “¿qué es una emoción?”. En la quinta sección revisaremos, en particular, la manera en la que una determinada concepción de las propiedades evaluativas parece estar condicionando varias de las teorías de las emociones que examinamos aquí.

2. ACLARACIONES INICIALES

2.1 *El enfoque*

A diferencia de las ciencias afectivas, cuyo interés fundamental consiste en “promover la medición y la experimentación con fines de predicción y explicación en una disciplina científica específica”¹, la filosofía de las emociones suele “tener como objetivo principal explicar o dar sentido a la experiencia humana de las emociones”² (SCARANTINO y DE SOUSA 2018). La experiencia humana de las emociones, no obstante, resulta ininteligible si nos aproximamos a ella al margen de las prácticas y los contextos en los que las emociones tienen lugar. Esto quiere decir que para analizar qué son las emociones hemos de partir de nuestro discurso cotidiano sobre ellas, el cual es claramente de tipo normativo: “no deberías sentirte culpable por esto”, “el comentario que le hizo fue muy ofensivo, normal que se enfadara”, “no entiendo cómo no derramó ni una lágrima durante el funeral”, etc. Siguiendo a teóricos de las emociones como Goldie (2000) o Mitchell (2021), el análisis que llevaremos a cabo aquí está guiado por una explicación de tipo *personal*, en contraste con las explicaciones impersonales que las ciencias afectivas nos ofrecen de las emociones. Las explicaciones de tipo personal están basadas en la idea de que cuando observamos que alguien responde emocionalmente a algo, asumimos que es *la persona* la que está respondiendo —y no su cuerpo o una serie de procesos causales y sistemas que son ajenos a la manera en la que la persona en cuestión aprehende la situación en la que se encuentra—.

Desde esta perspectiva, la comprensión de las respuestas emocionales es indisociable del hecho de que la persona que responde emocionalmente tiene un determinado punto de vista sobre la situación en la que se encuentra. Este es el tipo de explicación que más nos interesa si pretendemos dar cuenta del

¹ “promote measurement and experimentation for the purposes of prediction and explanation in a specific scientific discipline” (SCARANTINO y DE SOUSA 2018).

² “primary target making sense of the human experience of emotions” (*ibid*).

discurso normativo en el que se manifiesta nuestra experiencia de las emociones. Resulta evidente, por ejemplo, que cuando enseñamos a niñas y niños ante qué tipo de cosas han de sentir miedo (“no hay que asustarse de los payasos, pero sí de las personas desconocidas que se ofrecen a llevarnos en coche”) no nos estamos relacionando con amígdalas o sistemas límbicos, sino con personas que sienten miedo ante determinadas circunstancias. De lo contrario, esta práctica que regula nuestro miedo sería ininteligible.

Cualquier teoría que pierda de vista estos dos aspectos tan básicos —las respuestas emocionales (a) son indisociables del hecho de que la persona que responde emocionalmente tiene un determinado punto de vista y (b) están reguladas en el seno de un conjunto de prácticas y contextos (los cuales incluyen los hechos que descubre la investigación científica)— estará en contradicción con nuestra experiencia y discurso sobre las emociones. Por lo tanto, se tratará de una teoría que no contribuye a la comprensión del fenómeno que estamos investigando. A pesar de que la gran mayoría de filósofos de las emociones se propone ofrecer explicaciones compatibles con nuestro discurso normativo sobre emociones, más adelante veremos cómo hay teorías que no parecen cumplir realmente este objetivo.

2.2 *El fenómeno*

Una vez precisado el tipo de explicación que guiará nuestro análisis, es importante que restrinjamos el fenómeno al que normalmente hacen referencia los teóricos de las emociones. Aunque el dominio afectivo es muy amplio y requiere de un análisis mucho más detallado y profundo del que podemos llevar a cabo aquí, en el debate sobre emociones actual se suele diferenciar a las emociones de los estados de ánimo, rasgos de carácter, sentimientos y disposiciones emocionales³.

En la discusión contemporánea, el término *emoción* se reserva para estados *sentidos* y *episódicos* (DEONNA y TERONI 2012; TAPPOLET 2016; CARMAN 2018; PINEDA 2019; MITCHELL 2021)⁴. Ejemplos de este tipo son el miedo que experimento cuando veo un coche dirigirse descontroladamente en mi dirección, la irritación que siento cuando alguien no me toma en serio o la alegría que me produce encontrarme a una amiga por la calle. Un estado episódico

³ Para una discusión más detallada de estas diferencias ver Deonna y Teroni (2012) y Pineda (2019).

⁴ El aspecto sentido de las emociones —el hecho de que estas poseen una cierta fenomenología— se tratará en mayor profundidad en la siguiente sección, no obstante, resulta pertinente mencionarlo aquí para poder contrastar las emociones con otro tipo de fenómenos afectivos *no sentidos*.

es aquel que sólo es correcto atribuirle a alguien si está siendo instanciado por él en ese preciso momento. En el momento en que me encuentro a mi amiga por la calle, por ejemplo, es correcto atribuirme un episodio emocional de alegría, pero conforme pasa el encuentro, dejaría de ser correcto decir de mí que *estoy experimentando* alegría por el encuentro con mi amiga. Lo mismo ocurre con el caso del miedo: una vez pasado el susto, ya no siento miedo. Por descontado, dichos episodios pueden derivar en otro tipo de estados afectivos, como estados de ánimo (el encuentro con mi amiga puede ponerme de buen humor mientras que el susto con el coche puede dejarme intranquila) o rasgos de carácter (si, por el motivo que sea, experimento miedo de manera repetida, puedo acabar desarrollando un carácter temeroso). Puedo incluso volver a experimentar alegría o miedo si recuerdo lo vivido, en cuyo caso, sería correcto atribuirme una nueva emoción —un nuevo episodio de alegría o miedo—.

Los *estados de ánimo*, como las emociones, son estados episódicos y sentidos. Cuando decimos de alguien que se le ve contento, irritable o intranquilo estamos haciendo referencia a una experiencia que, en ese momento, alguien está atravesando y que se siente de una manera determinada. No obstante, los estados de ánimo se diferencian de las emociones por el tipo de objeto al que están dirigidos (GOLDIE 2000; MITCHELL 2021). Si nos fijamos en los ejemplos de emociones que hemos mencionado antes, podemos ver cómo, en todos los casos, las emociones están dirigidas (o son respuestas) a objetos particulares: el miedo está dirigido al coche; la irritación, a la persona que no me toma en serio; la alegría, a la amiga que me encuentro por casualidad. Como veremos en la siguiente sección con más detalle, las emociones son siempre *respuestas a objetos y situaciones concretas*. Los estados de ánimo, por su parte, también parecen ser respuestas a algo, sin embargo, su objeto es mucho más inespecífico que el de las emociones. Cuando estamos de buen humor o intranquilos, tanto nuestro buen humor como nuestra intranquilidad parecen dirigirse al mundo en general y no meramente a objetos particulares, como si a través de nuestro estado de ánimo el mundo en su conjunto se nos presentara de una manera nueva.

Esta capacidad para presentar el mundo bajo una determinada luz parece ser una característica que los estados de ánimo tienen en común con los *rasgos de carácter* (GOLDIE 2004; DEONNA y TERONI 2012). Una persona alegre percibe el mundo como un lugar amable y lleno de oportunidades, mientras que a una persona temerosa se le presenta como un lugar repleto de peligros. Los rasgos de carácter, sin embargo, se diferencian de los estados de ánimo (y de las emociones) en un aspecto fundamental: en lugar de estados episódicos, los rasgos de carácter son estados disposicionales. Un estado disposicional es aquel que se manifiesta exclusivamente ante ciertas circunstancias desencadenantes.

El hecho de que no se esté dando ninguna manifestación, no obstante, no implica que no sea correcto atribuirle tal estado a alguien (a diferencia de lo que ocurría con los estados episódicos). Es correcto atribuirle a alguien el rasgo de ser temeroso a pesar de que no se esté comportando continuamente de forma temerosa, pues basta con que dicho rasgo se manifieste en las circunstancias relevantes: si le proponemos a esta persona realizar alguna actividad de riesgo, no querrá participar; si hay un ascensor en el edificio, preferirá usar las escaleras; si tiene que planear un viaje, anticipará todo tipo de dificultades, etc. Al ser estados disposicionales, además, los rasgos de carácter *no son sentidos*, es decir, no son experiencias que atravesamos. Como mucho, lo que sentimos son algunas de sus manifestaciones (la angustia durante el viaje en avión, la tranquilidad de subir por las escaleras, etc.).

En la literatura sobre emociones suele pensarse que los *sentimientos* son otro tipo de estado afectivo disposicional cuyos ejemplos paradigmáticos son el amor y el odio. Los teóricos de las emociones coinciden en que amar u odiar a alguien no consiste en experimentar ciertos episodios de amor u odio hacia ese alguien (en cuyo caso estaríamos ante emociones) sino en un patrón de comportamiento y atención centrado en la persona amada u odiada (HELM 2009; GOLDIE 2010; NAAR 2013). En el caso del amor, por ejemplo, si la persona amada está presente, me resultará muy difícil desviar mi atención de ella; si se encuentra en apuros, querré ayudarle en lo que pueda; si surge la oportunidad de pasar tiempo con ella, no dudaré en aprovecharla, etc. Esto quiere decir que el amor y el odio son disposiciones que se manifiestan en circunstancias relevantes. Dichas manifestaciones pueden incluir, además, emociones muy variadas, por ejemplo, amar a una persona no es otra cosa que sentir alegría cuando se está cerca de ella, orgullo ante sus logros, miedo ante la posibilidad de que la relación con ella termine, etc. Como vemos, estas diferentes emociones tienen sentido en el contexto del sentimiento en cuestión. Del mismo modo que ocurría con los rasgos de carácter, el hecho de que un sentimiento no se esté manifestando en un momento particular no implica que dicho sentimiento haya desaparecido. Al tratarse de un estado disposicional, podemos decir de alguien que ama a otra persona a pesar de que, en ese momento, no esté pensando en ella o experimentando emociones hacia ella. Como decíamos, basta que su amor se manifieste en las circunstancias relevantes. El hecho de que el amor sea un estado disposicional implica, por otra parte, que se trata de un estado que *no es sentido*; lo que sentimos son sus diversas manifestaciones. A diferencia de los rasgos de carácter, no obstante, las manifestaciones de los sentimientos están unificadas de una manera muy particular: todas hacen referencia al mismo objeto —en el caso del amor, a la persona amada; en el caso del odio, a la persona odiada—.

Llegados a este punto, podría parecer que el conjunto de nuestra vida afectiva ha quedado cubierto por las diferentes categorías examinadas hasta ahora. No obstante, hay un sentido importante en el que los casos que presentábamos en el apartado correspondiente a las emociones admiten una interpretación diferente a la propuesta. En ocasiones, cuando decimos cosas como “estoy celosa de esa relación” o “estoy enfadada con mi madre” no estamos haciendo referencia a episodios emocionales concretos sino a nuestra disposición a sentir celos cada vez que, por ejemplo, vemos cómo se comporta nuestra pareja con cierta persona, le oímos hablar de ella, vemos que recibe un mensaje suyo, etc. o a nuestra disposición a sentir enfado cuando vemos a nuestra madre, recordamos la última conversación con ella, alguien nos la menciona, etc. Este tipo de estados afectivos reciben el nombre de *disposiciones emocionales* (DEONNA y TERONI 2012; TAPPOLET 2016; MITCHELL 2021) y se diferencian de los sentimientos en que las emociones que forman parte de sus manifestaciones son del mismo tipo: los celos disposicionales son una disposición a sentir celos en circunstancias determinadas mientras que el amor es una disposición a sentir alegría, orgullo, miedo, etc. en relación a la persona amada. Del mismo modo que los sentimientos, no obstante, y a diferencia de los rasgos de carácter, las manifestaciones de las disposiciones emocionales están todas dirigidas al mismo objeto. Los celos que siento cuando veo cómo mi pareja se comporta con cierta persona, le oigo hablar de ella o veo que recibe un mensaje suyo son, en todos los casos, celos que hacen referencia a su relación con esa otra persona.

El conjunto de estas distinciones constituye la visión más estándar dentro del debate actual sobre emociones. Hablar de emociones equivale, por tanto, a hablar de *episodios emocionales*. Este dato es importante porque dependiendo de a qué llamemos emoción, nuestra investigación se centrará en un fenómeno u otro. Hay dos excepciones importantes a esta visión estándar: Wollheim (1999) y Goldie (2000) prefieren reservar el término emoción para las llamadas disposiciones emocionales. Según Goldie, fijarnos exclusivamente en los episodios emocionales nos hace perder de vista el patrón o la narrativa que estructura y da sentido a tales episodios. Este tipo de patrones es lo que en el debate parece haber recibido el nombre de disposiciones emocionales, sin embargo, la idea de que este tipo de disposición sólo pueden incluir episodios del mismo tipo no parece correcta si atendemos a casos como el siguiente:

Estás celoso porque crees que se ha ido con otro. No puedes dormir: tu corazón y tu mente van a mil por hora durante toda la noche. Mientras te vistes por la mañana, no puedes evitar imaginártelos juntos, hablando y bromeando sobre ti

quizás, y eres incapaz de mantener tu mente en otra cosa. De camino al trabajo, ves a lo lejos a otra pareja y una de las dos personas se parece a ella, y prácticamente te desmayas, paralizado de terror. Más tarde, estás pendiente del trabajo durante un rato, y de repente, como un golpe en el cuerpo, ves en tu mesa algo de ella que vuelve a disparar tus sentimientos, y piensas: ‘Si no soy capaz de hablar con ella *ahora*, no sé qué voy a hacer’. Al minuto siguiente, tus celos dan un nuevo giro y esperas no volver a verla; suena el teléfono y la idea de que sea ella te llena de temor. Este conjunto de percepciones, pensamientos, sensaciones y cambios corporales son episodios dinámicamente relacionados de la misma emoción: el mismo estado de celos. Y estos elementos encajan como parte de una narrativa de esta parte de tu vida⁵

Cualquiera que viera *el conjunto* de los eventos narrados en este fragmento no diría que el protagonista está asustado, nervioso, enfadado o desesperado, sino celoso. Fijarnos de manera independiente en los diferentes episodios, no obstante, nos hace perder de vista cuál es su emoción y nos impide comprender lo que está ocurriendo en esta situación. Este tipo de consideraciones lleva a Goldie a plantearse hasta qué punto podemos alcanzar una buena comprensión de qué son las emociones a través de una teoría que se centre exclusivamente en episodios emocionales. Puede que una teoría de este tipo explique un parte del fenómeno, pero no está claro que lo que las emociones son pueda reducirse a meros episodios emocionales.

3. EMOCIONES: FENÓMENOS INTENCIONALES, EVALUATIVOS Y SENTIDOS

En la sección anterior hemos descrito el tipo de explicación personal desde el que vamos a aproximarnos a las emociones así como el tipo de fenómeno afectivo en el que suelen centrarse los teóricos de las emociones. Aunque veía-

⁵ “You are jealous because you think that she has run off with someone else. You cannot sleep: your heart and mind are racing all night. While you are getting dressed in the morning you cannot help imagining them together, talking and joking about you perhaps, and you are unable to keep your mind on anything else. On the way to work, you see another couple in the distance, one of whom looks just like her, and you practically faint, frozen to the spot in terror. Later in the day, you are preoccupied with work for a while, and then suddenly, like a blow to the body, you see on your desk something of hers which triggers your feelings again, and you think ‘If I’m not able to talk to her *now* then I don’t know what I’ll do’. The next minute your jealousy takes another turn, and you hope you never see her again; the telephone rings and the thought that it might be her fills you with dread. This complex of perceptions, thoughts, feelings, and bodily changes are dynamically related episodes of the same emotion—the same state of jealousy. And these elements fit in as part of a narrative of this part of your life” (GOLDIE 2000, 14).

mos que el consenso no es absoluto, el debate actual está organizado en torno a la idea de que las emociones son estados sentidos y episódicos. Antes de pasar a la revisión de las diferentes teorías de las emociones, describimos brevemente a continuación los tres aspectos característicos de las emociones que cualquier teoría satisfactoria habría de ser capaz de explicar, a saber: su aspecto intencional, evaluativo y sentido.

3.1 El aspecto intencional de las emociones

Un *estado intencional* es aquel que está dirigido a o versa sobre un objeto, situación, evento o estado de cosas —ya sea pasado, presente, futuro, actual, posible, imposible o ficticio—. Ejemplos de este tipo de estado son mi percepción visual de un árbol o mi recuerdo de las pasadas vacaciones. En el primer caso, mi percepción visual está dirigida al árbol (un objeto actual) y, en el segundo, mi recuerdo está dirigido a las pasadas vacaciones (un evento pasado). Técnicamente, decimos que aquello a lo que los estados intencionales están dirigidos constituyen sus objetos intencionales: el árbol constituye el objeto intencional de mi percepción visual mientras que las pasadas vacaciones constituyen el objeto intencional de mi recuerdo.

La manera más extendida de entender el modo en la que los estados intencionales están dirigidos a sus objetos es pensar que tales estados *representan* sus objetos intencionales. Decimos, así, que mi percepción visual no sólo está dirigida hacia el árbol sino que lo representa de una manera determinada. Del mismo modo, mi creencia de que estoy escribiendo representa el estado de cosas compuesto por mi estar escribiendo (y, al tratarse de una creencia, lo representa como algo verdadero) y mi deseo de que termine el invierno representa el estado de cosas en el que el invierno ha terminado (y, al tratarse de un deseo, lo representa como un estado de cosas que debería darse). La característica más importante de la noción de representación es que esta conlleva la posibilidad del error, esto es, la idea de que una representación puede ser mejor o peor. Así, una percepción visual puede ser verídica o no verídica; una creencia, verdadera o falsa; un recuerdo, preciso o impreciso, etc.

Tal y como sugeríamos en la sección anterior, las emociones son siempre respuestas a objetos y situaciones concretas. Esto quiere decir que las emociones son, claramente, estados intencionales. Cuando siento orgullo ante el cuadro que he pintado, mi orgullo está dirigido al cuadro y lo representa de una manera determinada. El cuadro, por su parte, constituye el objeto intencional de mi orgullo. En el caso en el que siento alegría al encontrarme a una amiga por la calle, podemos ver cómo, por una parte, mi alegría está dirigida a mi amiga y cómo, por otra, mi amiga constituye el objeto intencional de

mi emoción. El hecho de que las emociones sean estados intencionales implica que, del mismo modo que percepciones y creencias, estas están sujetas a estándares de corrección y pueden ser evaluadas como respuestas apropiadas o inapropiadas. Si mi cuadro podría haber sido pintado por un niño de tres años, parece evidente que mi cuadro no merece orgullo y que, por tanto, mi emoción no es una respuesta proporcional a su objeto. Si mi amiga es una persona a la que aprecio, sentir alegría al verla es una respuesta apropiada a la situación.

3.2 El aspecto evaluativo de las emociones

Como acabamos de mencionar, las emociones representan su objeto intencional *de una manera determinada*. El orgullo que siento hacia el cuadro que he pintado representa el cuadro de una manera muy distinta a cómo, por ejemplo, lo representaría mi vergüenza si pensara que se trata de un cuadro que podría haber pintado un niño de tres años. Una de las ideas más asentadas en el debate actual es que las emociones siempre representan o aprehenden el objeto al que están dirigidas bajo un determinado *aspecto evaluativo* (KENNY 1963; SOLOMON 1993; NUSSBAUM 2001; DEONNA y TERONI 2012; TAPPOLET 2016; CARMAN 2018; CAMPEGGIANI 2021; MITCHELL 2021): el orgullo representa su objeto como algo encomiable; la vergüenza, como algo vergonzoso; el miedo, como algo peligroso; la tristeza, como algo que supone una pérdida, la culpa, como algo reprochable, etc. Ser encomiable, vergonzoso, peligroso u reprochable son diferentes tipos de *propiedades evaluativas* que nos permiten, en principio, individuar los distintos tipos de emoción. Podemos, así, identificar el orgullo que siento por mi cuadro, mi hija y el trabajo bien hecho como diferentes instancias del mismo tipo de emoción en virtud de que son todas respuestas a algo que se representa como encomiable. *A cada tipo de emoción*, por lo tanto, le corresponde un tipo de propiedad evaluativa.

Aunque la idea de que las emociones están estrechamente conectadas con las distintas propiedades evaluativas está muy extendida en el debate sobre emociones actual, hay teorías que no recogen este aspecto evaluativo (JAMES 1884; GORDON 1987; GREEN 1992). No obstante, no está claro cómo podríamos dar cuenta del discurso normativo de emociones si prescindiéramos de dicho aspecto: como veímos, las emociones pueden ser evaluadas como respuestas apropiadas o inapropiadas, es decir, están sujetas a estándares de corrección o propiedad. Estos estándares, por su parte, parecen inseparables de las propiedades evaluativas que acabamos de introducir: decir del cuadro que no merece orgullo —y que, por tanto, sentir orgullo por él es una respuesta desproporcional— equivale a decir que el cuadro no es *encomiable*. Igualmen-

te, decir que sentir enfado ante un insulto es una respuesta apropiada equivale a decir que el insulto es realmente *ofensivo*⁶. Si no pudiéramos hacer referencia a las diferentes propiedades evaluativas, no está claro cómo podríamos corregir o censurar las respuestas emocionales de alguien. Por este motivo, y en consonancia con una gran parte de los teóricos de las emociones actuales, consideraremos el aspecto evaluativo de las emociones como un elemento a explicar por una teoría satisfactoria de las emociones. Otro motivo importante es que, como señala Campeggiani, parece que una experiencia emocional sólo resulta inteligible si tenemos en cuenta su aspecto evaluativo (2021, 110). A fin de ilustrar esta idea, Campeggiani introduce la siguiente cita a Kenny:

[S]i un hombre dice que siente remordimientos por las acciones de alguien que no tiene nada que ver con él, o tiene envidia de sus propios vicios, *no podemos entenderle* (...) es posible estar orgulloso de un vicio o de un crimen o de un defecto, si uno puede representárselo a sí mismo como una virtud o un logro o una ventaja (...) Lo que no es posible es estar agradecido u orgulloso de algo que uno considera completamente negativo. También es posible tener envidia de los frutos de un árbol propio; pero sólo si uno cree erróneamente que el terreno en el que está el árbol es parte de la propiedad del vecino (...) Lo que no es posible es envidiar algo que uno cree que le pertenece a uno mismo⁷

3.3 *El aspecto sentido de las emociones*

Como indicábamos en la sección anterior, existe un cierto consenso en torno a la idea de que las emociones son estados episódicos, los cuales se caracterizan por ser *sentidos de una manera determinada*, esto es, por poseer una

⁶ Las propiedades evaluativas suelen recibir el nombre de “objetos formales” de las emociones (KENNY 1963). Hay quien piensa que el objeto intencional de las emociones está compuesto por dos tipos de objeto: un objeto particular (el objeto al que está dirigido la emoción) y un objeto formal (la propiedad evaluativa en cuestión) (DEONNA y TERONI 2012; TAPPOLET 2016; CARMAN 2018; PINEDA 2019). No obstante, el hecho de que los objetos formales proporcionen el criterio normativo para distinguir entre los diferentes tipo de emoción no implica que la emoción represente o esté dirigida a tales objetos formales (MITCHELL 2021). El objeto formal de una creencia, por ejemplo, es la verdad (una creencia representa su objeto *como verdadero*), pero no diríamos que la creencia representa un determinado estado de cosas *y, además*, la verdad.

⁷ “if a man says that he feels remorse for the actions of someone quite unconnected with him, or is envious of his own vices, *we are at a loss to understand him* (...) [it] is possible to be proud of a vice or a crime or a defect, if one can represent it to oneself as a virtue or an achievement or an advantage (...) What is not possible is to be grateful for, or proud of, something which one regards as an evil unmixed with good. Again, it is possible to be envious of one's own fruit trees; but only if one mistakenly believes that the land on which they stand is part of one's neighbour's property (...) What is not possible is to envy something which one believes to belong to oneself” (KENNY 1963, 134-5, énfasis añadido).

determinada fenomenología. La idea que parecen compartir la mayoría de los teóricos de las emociones es que si nos enteráramos, por ejemplo, de que alguien está avergonzado por algo que ha hecho, no sólo sabríamos que su vergüenza está dirigida a lo que ha hecho y que lo representa como algo vergonzoso, sino que sabríamos también *cómo se siente* esta persona (DEONNA y TERONI 2012) —en qué consiste su *experiencia* emocional, el tipo de sensaciones que está atravesando, etc.—. Desde esta perspectiva, la vergüenza se siente de un modo particular que es muy distinto a cómo se sienten, por ejemplo, los celos o la esperanza⁸.

Una de las cuestiones que más ha preocupado a los participantes en este debate es cómo dar cuenta de la cualidad sentida de las emociones. Por una parte, hay un importante número de filósofos que explican la fenomenología de las emociones en términos de la *percepción de cambios corporales*. Sentir una emoción es equivalente, según esta visión, a percibir desde dentro una serie de cambios fisiológicos (JAMES 1884; DEONNA y TERONI 2012). Este tipo de cambios incluyen cambios faciales y vocales (sonrisa, ceño fruncido, voz aguda, etc.), cambios musculares (postura erguida, encogimiento de hombros, parálisis, etc.), cambios del sistema nervioso autónomo (aceleración del ritmo cardíaco, de la respiración, sudoración, etc.) y cambios del sistema endocrino (subidón de adrenalina). Desde esta perspectiva, la cualidad sentida de la vergüenza, por ejemplo, se agota en mi percepción interna de cómo me sonrojo, se me acelera el corazón, mi cabeza se agacha, etc. No es difícil ver que detrás de esta concepción de la fenomenología de las emociones hay una importante vocación empírica, pues los diferentes tipos de cambios fisiológicos mencionados son todos *medibles*.

Hay quien piensa, por otra parte, que el modo en el que las emociones se sienten no puede ser reducido a la mera percepción de cambios corporales o, dicho de otro modo, a una fenomenología de tipo *corporal*, pues parecería que sentir cambios corporales no equivale a sentir emociones. Cuando, por ejemplo, al sentir vergüenza decimos que no queremos sentirnos de este modo, no estamos queriendo decir que no queremos notar cómo se nos enrojece la cara o cómo se nos acelera el corazón. Según esta visión, la fenomenología de la vergüenza parece constituir, más bien, una forma de sentirse muy compleja (y poco agradable) que perdemos de vista si intentamos fragmentarla en dife-

⁸ Las percepciones y los dolores, por ejemplo, son también estados sentidos o fenomenológicos. Los dolores, no obstante, no pueden ser evaluados como respuestas proporcionales o desproporcionales de la misma manera que las emociones. Un dolor de muelas, por ejemplo, no es una respuesta a una situación que el sujeto registra desde su punto de vista evaluativo. Por este motivo, tendemos a pensar que los dolores no son estados intencionales. No obstante, para una concepción del dolor como estado intencional ver Helm (2002).

rentes cambios corporales. Una alternativa posible es considerar que la fenomenología de la vergüenza no consiste meramente en sentir cómo me sonrojo, sino que, cuando siento vergüenza, *me siento indigna* (DEIGH 2010, 25). El tipo de fenomenología que se intenta explicar desde esta perspectiva podría calificarse de fenomenología *vivencial* (VENDRELL 2018): en la humillación, me siento pisoteada; en el orgullo, engrandecida; en la tristeza, desamparada, etc. Las sensaciones que acompañan a cada tipo de emoción, no obstante, parecen muy difíciles de describir sin hacer uso de un lenguaje metafórico. Esto podría ser precisamente lo que Goldie tiene en mente cuando señala que “si has experimentado la emoción, entonces sabes muy bien cómo pueden ser este tipo de sentires o sensaciones, y no necesitas que yo te lo diga; si no has experimentado la emoción y quieres hacerte al menos una idea de cómo se siente, entonces (...) te sugiero que leas una buena novela”⁹ (2000, 19).

4. TEORÍAS DE LAS EMOCIONES

De todo lo visto hasta ahora se desprende que una buena teoría de las emociones habrá de explicar el hecho de que las emociones están intencionalmente dirigidas a objetos y situaciones, que representan su objeto intencional bajo un determinado aspecto evaluativo y que poseen una importante cualidad sentida. Además, el tipo de explicación ofrecida tendrá que tener en cuenta que quien responde emocionalmente tiene un determinado punto de vista sobre la situación en la que se encuentra y que es *ella*, y no su cuerpo, quien responde emocionalmente a tal situación. Como veremos, el conjunto de teorías que revisaremos a continuación tiene por objetivo dar cuenta de las emociones entendidas como episodios emocionales.

4.1 Teorías cognitivistas

Vamos a comenzar nuestro análisis haciendo referencia a las teorías de las emociones de corte más cognitivista: la teoría creencia-deseo y la teoría jacialista. A pesar de que, a día de hoy, suele pensarse que estas teorías están ampliamente superadas, resulta iluminador prestar atención a sus motivaciones y dificultades porque nos permiten entender a qué tipo de problemas intentan dar respuesta las teorías sucesivas. Cronológicamente, deberíamos empezar

⁹ “if you have experienced the emotion, then you know very well what these sorts of feeling can be like, and you do not need me to tell you; if you have not experienced the emotion and want to get at least some idea of what it feels like (...) I suggest you read a good novel” (GOLDIE 2000, 19).

examinando la teoría jamesiana, pero por motivos teóricos, la reservaremos para más adelante.

4.1.1 La teoría creencia-deseo

Los teóricos de la creencia-deseo consideran que el aspecto más importante de las emociones es su aspecto intencional. Para filósofos como Gordon (1987) y Green (1992), cuando nos planteamos la pregunta “¿qué son las emociones?”, no estamos tratando de comprender un fenómeno intencional *sui generis*, pues, según ellos, las emociones no son otra cosa que *combinaciones de creencias y deseos*¹⁰. La idea fundamental es que la tristeza que alguien puede sentir ante la grave enfermedad de un amigo consiste en su creencia de que su amigo está muy enfermo y su deseo de poder pasar más tiempo junto a él. Asimismo, el miedo ante un perro consiste en la creencia de que el perro está gruñendo y dispuesto a lanzarse sobre nosotros y el deseo de salir corriendo. Para poder dar cuenta de emociones dirigidas a estados de cosas que no son realizables —porque están dirigidas al pasado o situaciones imaginarias— y en cuyo caso no podemos atribuir un deseo al sujeto, Gordon sostiene que las emociones pueden identificarse con la combinación de una creencia y un estado motivacional diferente, por ejemplo, un anhelo. Así, mi arrepentimiento por haber mentido a una amiga consiste en mi creencia de que no le he dicho la verdad y mi anhelo de haber actuado de otra manera.

Un aspecto importante de la teoría creencia-deseo es que esta identifica la parte sentida de las emociones con la sensación de cambios corporales (esto es, con una fenomenología de tipo corporal) y considera que esta no es un elemento constitutivo de las emociones. La idea es que esta parte sentida es un efecto *causal* de las emociones, por lo que no podemos identificar las emociones con cómo se sienten. Por ejemplo, sentir cómo se me pone la cara roja de vergüenza no es parte de la vergüenza o, más precisamente, no *es* la vergüenza, sino mi percepción de un síntoma causado por la propia emoción. Desde esta perspectiva parece seguirse que nos equivocamos cuando decimos que *sentimos* vergüenza: siendo precisos, lo que sentimos son los efectos causales de la vergüenza (PINEDA 2019).

El tratamiento que la teoría creencia-deseo ofrece de la fenomenología de las emociones parece problemático por dos motivos. El primero ha sido ya expuesto en la sección anterior: tal y como señalan los críticos de la fenomenología corporal, no está claro que sentir cambios corporales equivalga a sentir emociones. El segundo es que, al relegar el aspecto sentido de las emociones a un segundo plano y sostener que estas no son sentidas, la teoría creencia-deseo

¹⁰ En una línea similar, Davidson ha defendido en su ensayo “La teoría cognitiva del orgullo de Hume” una teoría proposicional de las emociones (1980).

parece estar renunciando a uno de los aspectos más importantes de las emociones que nos permiten, precisamente, distinguirlas de otro tipo de fenómenos mentales. Si las emociones no son más que combinaciones de creencias y deseos, no podríamos diferenciar entre casos en los que un sujeto tiene una emoción (constituida por una determinada combinación de creencias y deseos) y casos en los que un sujeto tiene la misma combinación de creencias y deseos pero no una emoción. Imaginemos que alguien ofende gravemente mi honor. Puedo ponerme furiosa y decidir vengarme (y mi emoción no sería otra cosa que la combinación de la creencia de que mi honor ha sido mancillado y el deseo de que mi reputación permanezca intacta) o no verme especialmente afectada pero tomar represalias simplemente porque no responder a la ofensa sería percibido como una muestra de debilidad (en cuyo caso, creo que mi honor ha sido mancillado y deseo que mi reputación permanezca intacta). Si decimos, como hace la teoría creencia-deseo, que las emociones no son otra cosa que combinaciones de creencias y deseos no podemos ver cuál es la diferencia entre estos dos casos. Parece, no obstante, que estos casos son diferentes en un sentido relevante. Decir que la diferencia radica en que en el primer caso, pero no en el segundo, el sujeto sentirá una serie de cambios corporales no parece una explicación plausible de la naturaleza de las emociones (GOLDIE 2000). Sobre todo, teniendo en cuenta que hay emociones que no parecen comportar cambios corporales (como, por ejemplo, el orgullo).

Una de las objeciones más típicas a la teoría creencia-deseo es que, al identificar las emociones con creencias y deseos, dicha teoría atribuye a las emociones un contenido de tipo proposicional, lo cual restringe la capacidad para tener emociones a la capacidad de tener lenguaje (DEIGH 2010). Resulta evidente, no obstante, que tanto animales como niños de corta edad, a pesar de carecer de lenguaje, tienen emociones, por lo tanto, esta teoría no puede ser correcta. Esta objeción suele presentarse como una dificultad insalvable para la teoría creencia-deseo, sin embargo, el presupuesto fundamental en el que esta se basa —que la capacidad para tener creencias y deseos depende de la posesión del lenguaje— es un presupuesto filosóficamente controvertido. Tal y como ha argumentado Dennett (1995), podemos comprender, explicar y predecir el comportamiento de los animales si nos aproximamos a él como un comportamiento guiado por ciertos deseos dadas ciertas creencias:

El mapache quiere la comida que hay en la caja-trampa, pero sabe que no debe entrar en una trampa potencial donde no puede ver la salida. Por eso hay que poner dos puertas abiertas en la trampa, para que el animal se atreva a entrar por la primera, planeando salir por la segunda si hay algún problema. Resultará muy difícil conseguir que un mapache entre en una trampa que no tenga una

aparente “salida de emergencia” que se cierre junto con la puerta de entrada. Entiendo que este estilo de explicación y predicción es incontrovertidamente valioso: funciona, y funciona porque los mapaches (por ejemplo) son así de listos. Este hecho basta, dado lo que entiendo por “creencia”, para demostrar que los mapaches tienen creencias —y deseos, por supuesto—¹¹

Por lo tanto, y a diferencia de lo que suele pensarse, no está tan claro que la teoría creencia-deseo sea incapaz de explicar las emociones de los animales y los niños de corta edad. La objeción que sí parece poner en un aprieto a esta teoría es la que afirma que apelar a combinaciones de creencias y deseos no nos permite individuar los diferentes tipos de emoción (DEONNA y TERONI 2012). Que llamemos miedo a lo que sentimos ante un perro (creo que el perro me está gruñendo y deseo salir corriendo), ante la desaparición de un documento importante (creo que el documento no está en la carpeta en la que lo guardé y deseo conservarlo) y la posibilidad de perder el tren (creo que no voy a llegar a la estación a tiempo y deseo realizar el viaje) ha de ser explicado en virtud de algo común a estas tres experiencias. Sin embargo, no hay nada en las diferentes combinaciones de creencias y deseos que explique que son todas instancias de miedo y no de cualquier otra emoción.

Un defensor de la teoría creencia-deseo podría argumentar que el tipo de deseos que forman parte de la combinación creencia-deseo constitutiva de las emociones son deseos con contenido abierto. Lo que distinguiría, entonces, al miedo sería el deseo de evitar cualquier tipo de daño; a la envidia, el deseo de tener algo de lo que carecemos; la admiración, el deseo de estar en contacto con objetos estéticos, etc. El problema de esta estrategia es que al postular deseos con contenido abierto se pierde el vínculo interno entre el contenido de las creencias y los deseos. La creencia de que el perro está gruñendo (va a atacarme) y el deseo de salir corriendo (evitar que me ataque) están conceptualmente conectados. Sin embargo, un sujeto que tenga el deseo con contenido abierto de evitar cualquier tipo de daño no tendrá una emoción de miedo hasta que capte —a través de un estado mental adicional— por qué su creencia de que el perro le está gruñendo es relevante en relación al contenido de ese deseo (DEONNA y TERONI 2012).

¹¹ “The raccoon wants the food in the box-trap, but knows better than to walk into a potential trap where it cannot see its way out. That is why you have to put two open doors on the trap-so that the animal will dare to enter the first, planning to leave by the second if there is any trouble. You will have a hard time getting a raccoon to enter a trap that does not have an apparent “emergency exit” that closes along with the entrance. I take it that this style of explanation and prediction is un controversially valuable: it works, and it works because raccoons (for instance) are that smart. That fact suffices, given what I mean by “belief,” to show that raccoons have beliefs-and desires, of course” (DENNETT 1995, 112).

En conclusión, la teoría creencia-deseo parece centrarse exclusivamente en el aspecto intencional de las emociones dejando de lado su aspecto evaluativo y fenomenológico. Su explicación del aspecto intencional, además, no parece ser muy plausible dadas las objeciones planteadas a la reducción de las emociones a creencias y deseos.

4.1.3 La teoría juicialista

El rótulo “teoría juicialista” hace referencia a una familia de teorías de las emociones cuya tesis común es que las emociones son *un tipo de juicio evaluativo* sobre objetos que nos conciernen (SOLOMON 1993; NUSSBAUM 2001). Siendo precisos, la teoría juicialista ha ido evolucionando a lo largo de los años y la analogía con los juicios ha adoptado diferentes formas. No obstante, como primera aproximación, nos centraremos en la identificación de las emociones con juicios evaluativos.

Para la teoría juicialista, las emociones son estados en los que juzgamos la relevancia que un objeto o situación tiene en relación a nuestro bienestar e intereses. La relación con nuestro bienestar se expresa a través de *conceptos de tipo evaluativo* (como por ejemplo, “vergonzoso”, “terrorífico”, “ilusionante”, etc.). Así, según esta teoría, el miedo al perro consiste en juzgar que el perro es peligroso mientras que estar triste por la enfermedad de un amigo consiste en juzgar que su posible muerte constituye una importante pérdida. Del mismo modo que la teoría creencia-deseo, la teoría juicialista considera que la intencionalidad es un aspecto esencial de las emociones, no obstante y a diferencia de la teoría anterior, el juicialismo incluye el aspecto evaluativo de las emociones como elemento ineliminable de su análisis. Esto supone una importante ventaja sobre la teoría creencia-deseo, pues al incluir conceptos evaluativos en los juicios que constituyen las emociones, los juicialistas permiten la individuación de los diferentes tipos de emoción (DEIGH 1994; ROBERTS 2003). Podemos clasificar el miedo al perro, el miedo a perder un documento importante y el miedo a no llegar a tiempo al tren como instancias de la emoción de miedo (y no de otra emoción) dado que, en los tres casos, el sujeto juzga que el objeto o situación en cuestión constituye una amenaza para su bienestar o intereses.

A pesar de que la inclusión del aspecto evaluativo constituye un avance en nuestra comprensión de la naturaleza de las emociones, la teoría juicialista relega a un segundo plano el aspecto sentido de las emociones. En línea con la teoría anterior, esta considera que los cambios corporales característicos de las emociones son *causados* por las propias emociones y que la fenomenología de la experiencia emocional se reduce a la sensación de tales cambios (los

cuales pueden incluir una serie de disposiciones a la acción). Como hemos visto antes, sostener que el aspecto sentido de las emociones no es una parte constituyente de las mismas y reducir dicho aspecto sentido a su dimensión puramente corporal no parece constituir una buena explicación de la fenomenología de las emociones (VENDRELL 2018).

Las críticas a la teoría juicialista, no obstante, se han centrado principalmente en las graves consecuencias que se derivan de la identificación de las emociones con juicios evaluativos. En primer lugar, suele pensarse que de esta identificación se sigue que, para tener emociones, un sujeto ha de poseer ciertos conceptos evaluativos, lo cual implicaría que los animales no humanos y los niños de corta edad son incapaces de sentir emociones (DEIGH 2010). Ante esta objeción, sin embargo, parece que podemos aplicar la respuesta que ofrece Dennett a la objeción de que los animales no tienen creencias y deseos: cuando vemos que el mapache sólo entra en la caja-trampa si hay una puerta de salida, podemos explicar y predecir su conducta atribuyéndole juicios sobre el nivel de amenaza que supone la situación en la que se encuentra. De nuevo, este tipo de explicación “funciona, y funciona porque los mapaches (...) son así de listos” (1995, 112).

La objeción más importante a la teoría juicialista es que no parece que tener un determinado juicio evaluativo sea una condición necesaria ni suficiente para sentir una emoción (D'ARMS y JACOBSON 2003; DEIGH 2010; DEONNA y TERONI 2012; TAPPOLET 2016; PINEDA 2019). Por una parte, podemos tener una emoción en ausencia del juicio evaluativo correspondiente: aunque juzguemos que un puente de cristal es extremadamente seguro y que, por tanto, no corremos ningún riesgo al cruzarlo, resulta prácticamente imposible no experimentar terror a caer al vacío que vemos a través del cristal. Por otra parte, podemos juzgar que una situación es, por ejemplo, triste y no sentir la emoción de tristeza correspondiente. En ocasiones, ante la muerte de un ser querido —la cual evaluamos con una importante pérdida para nosotros— somos incapaces de experimentar tristeza y de llorar, por ejemplo.

Las emociones que no se alinean con los juicios evaluativos del sujeto reciben el nombre de emociones *recalcitrantes*. Los ejemplos que acabamos de introducir muestran cómo este es un fenómeno bastante común. La explicación que la teoría juicialista ofrece del mismo, no obstante, ha sido objeto de importante debate. Al considerar que las emociones son juicios evaluativos, cuando el sujeto siente emociones recalcitrantes, lo que está ocurriendo, según el juicialista, es que está incurriendo en una contradicción (juzga, a la vez, que el puente es y no es peligroso). La pregunta es si este es realmente el tipo de irracionalidad en la que el sujeto incurre cuando su emoción no

se corresponde con su juicio evaluativo (BRADY 2009; DEIGH 2010; HELM 2015; GRZANKOWSKI 2017). No parece haber ninguna incoherencia en que el sujeto juzgue que el puente no es peligroso a pesar de que cruzarlo le aterrorice o, al menos, no el mismo tipo de incoherencia que existe cuando sostengamos que llueve y no llueve.

El problema que suponen las emociones recalcitrantes y la posibilidad de experimentar emociones sin juicios evaluativos (y viceversa) sugieren que la identificación entre juicios y emociones no puede ser del todo correcta. Algunos autores han intentado esquivar estas dificultades precisando que las emociones constituyen un tipo especial de juicio evaluativo. Según Nussbaum, las emociones son un tipo de juicio que pueden mantenerse a pesar de que estén en contradicción con otro de nuestros juicios (2001). Solomon, en cambio, sostiene que las emociones han de ser entendidas como “juicios del cuerpo” (2004). El problema principal que se ha atribuido a esta estrategia es que el tipo de juicios evaluativos postulados por la teoría parecen completamente *ad hoc*: se expande el concepto de juicio con el único objetivo de dar cuenta de las objeciones recibidas. La identificación de las emociones con juicios evaluativos se vuelve, así, trivialmente verdadera y no nos proporciona ningún tipo de comprensión relevante de la naturaleza de las emociones (SCARANTINO 2010).

Llegados a este punto, podemos preguntarnos si las críticas a la teoría juicialista, las cuales se presentan como devastadoras en el debate actual, son del todo justas con esta teoría. En particular, da la impresión de que las críticas a la teoría juicialista presuponen una concepción muy intelectualista de los juicios (y de las creencias, en el caso de los críticos de la teoría creencia-deseo): juzgar que algo es de determinada manera implica el uso de proposiciones y la posesión del lenguaje. Normalmente, se acusa al juicialismo de sobreintelectualizar a las emociones, sin embargo, esto sólo es evidente desde la perspectiva de quien sobreintelectualiza los juicios, pues la acusación de estar introduciendo un tipo de juicio *ad hoc* presupone una noción de juicio muy concreta como paradigma. Considerar que las emociones son un tipo de juicio podría abrirlas, tal vez, a la posibilidad de redefinir nuestro concepto de juicio si descubrimos que se trata de un concepto que sólo nos permite explicar un conjunto de casos fuertemente intelectualizados.

En cualquier caso, la teoría juicialista ha sido ampliamente rechazada en los últimos años. A pesar de que, a diferencia de la teoría creencia-deseo, el juicialismo ofrece una explicación del aspecto intencional y evaluativo de las emociones (dejando de lado, no obstante, su aspecto fenomenológico) se suele pensar, como hemos visto, que la asimilación de las emociones a juicios evaluativos que propone esta teoría no es convincente.

4.2 La teoría perceptiva

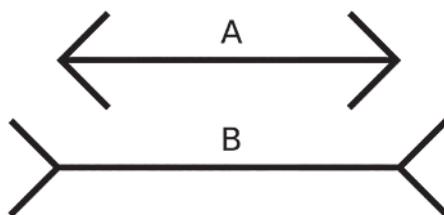
La teoría perceptiva se toma muy en serio las objeciones presentadas contra la teoría juicialista. De hecho, la mejor manera de aproximarse a su propuesta principal es considerándola una respuesta a los problemas generados por el juicialismo. Del mismo modo que las teorías que hemos examinado hasta ahora, la teoría perceptiva recurre a otros estados mentales para dar cuenta de las emociones. En este caso, la naturaleza de las emociones se intenta explicar a partir del modelo de la percepción (DE SOUSA 1987; ROBERTS 2003; TAPPOLET 2000). Aunque existen diferentes teorías perceptivas, la tesis común a todas ellas es que las emociones son *percepciones de valores o propiedades evaluativas*.

La idea fundamental es que, del mismo modo que percibimos el azul de una pelota a través de una experiencia visual, percibimos la peligrosidad del perro a través de una experiencia emocional de miedo. La experiencia perceptiva consiste en *la percepción del azul* de la pelota, y la experiencia emocional de miedo consiste en *la percepción de la peligrosidad* del perro. Las percepciones sensoriales, por una parte, nos pondrían en contacto con una serie de propiedades, en principio, *no evaluativas* del mundo: color, tamaño, distancia, forma, etc. mientras que las emociones; por otra, nos pondrían en contacto con propiedades *evaluativas* del mundo: la propiedad de ser peligroso, ofensivo, admirable, envidiable, etc.

Los defensores de esta teoría han ofrecido varios motivos por los que las percepciones proporcionan un modelo más adecuado para explicar las emociones. El primero de ellos es que las percepciones son estados fenomenológicos cuya cualidad sentida no puede ser captada por referencia a un juicio (juzgar que la pelota es azul carece de la fenomenología propia de percibir que la pelota es azul). Del mismo modo, sostiene la teoría perceptiva, la cualidad sentida de las emociones desaparece si las reducimos a su correspondiente juicio evaluativo (juzgar que el perro es peligroso carece de la fenomenología propia de temer al perro peligroso).

El segundo motivo es que las percepciones son, en principio, “una forma de cognición más primitiva” (DEIGH 2010, 28), es decir, no requieren que el sujeto posea los conceptos involucrados en el contenido de su experiencia perceptiva. Según esta teoría, un animal o un niño de corta edad puede percibir una pelota azul sin poseer el concepto “pelota” o “azul”. Si asumimos que las emociones son análogas a las percepciones en este sentido, parece que sorteamos una de las objeciones más típicas en contra de las teorías cognitivistas. Desde esta perspectiva, un niño puede percibir la peligrosidad de un animal, por ejemplo, sin poseer el concepto de lo peligroso del mismo modo que puede percibir visualmente el azul de una pelota sin poseer aún el concepto de lo azul. Consider-

rar que las emociones no tienen por qué involucrar un contenido conceptual ha llevado a los teóricos de la percepción a afirmar que su teoría puede responder a otra de las objeciones importantes a la teoría juicialista. Si recordamos, los casos en los que las emociones se encuentran en contradicción con los juicios evaluativos del sujeto suponen un problema para la teoría juicialista. Dado que dicha teoría identifica las emociones con juicios evaluativos, el juicialista se ve obligado a decir que en tales casos el sujeto tiene juicios contradictorios pero, como comentábamos antes, este no parece un análisis plausible del fenómeno de las emociones recalcitrantes. Este tipo de casos, defiende la teoría perceptiva, pueden explicarse fácilmente en base al modelo de las ilusiones perceptivas. El ejemplo que suele introducirse es el de la ilusión Müller-Lyer:



Si nos fijamos bien, las dos líneas horizontales son exactamente de la misma longitud, no obstante, debido a las líneas diagonales de sus extremos, nuestra impresión visual es que la línea inferior (B) es más larga que la superior (A). Aunque juzguemos que A y B tienen la misma longitud, percibimos B como más larga que A. La idea es que el contenido de la percepción puede ser “cognitivamente impenetrable” (GOLDIE 2000, 76) sin que ello implique que el sujeto se contradiga y sea, por tanto, incoherente. Según la teoría perceptiva, las emociones son análogas a las percepciones en este sentido también. Del mismo modo que no pensamos que seamos incoherentes por ver las dos líneas como de diferente longitud a pesar de que sabemos que tienen la misma, no parece que seamos incoherentes tampoco, por ejemplo, en el caso en el que tenemos miedo de una pequeña araña a pesar de que sabemos que no es peligrosa (TAPPOLET 2016).

Por último, resulta evidente que tanto las emociones como las percepciones son fenómenos intencionales, por lo que la teoría perceptiva puede retener la gran ventaja de las teorías cognitivistas. Además de poseer la capacidad para estar dirigidas a distintos objetos, parece que las percepciones, al igual que las emociones, nos proporcionan información sobre el mundo de una manera particular. Según un importante número de filósofos, las emociones, como las percepciones, nos proporcionan razones revisables para la formación de juicios evaluativos (GREENSPAN 2011; TAPPOLET 2016; CARMAN 2018).

Estos presupuestos tan básicos de la teoría perceptiva son, no obstante, cuestiones muy controvertidas en filosofía —de hecho, no constituyen en absoluto la visión más extendida de cómo funciona la percepción—. La idea de que esta consiste en el *registro pasivo* de lo que nos rodea y que no tiene por qué involucrar nuestras capacidades conceptuales es una idea muy problemática, principalmente, porque no está nada claro cómo partiendo de impresiones brutas podríamos formarnos las creencias y los juicios correspondientes a nuestras percepciones (McDOWELL 1994). Aun así, tanto la teoría perceptiva como las críticas a esta están construidas en base a este modelo¹². Una consecuencia importante de esto es que el tipo de objeciones que se han planteado a esta teoría no parecen arrojar mucha luz sobre qué son las emociones a menos que aceptemos la visión de la percepción propuesta. A continuación, examinamos brevemente estas críticas.

Las objeciones a la teoría perceptiva intentan mostrar que las diferencias entre emociones y percepciones son mucho más significativas que sus similitudes. La primera gran diferencia es que las percepciones son probablemente el único estado mental que no necesita que ningún otro estado mental le proporcione acceso o “contacto” con su objeto intencional (DEONNA y TERONI 2012; MITCHELL 2021). Esta parece ser, sin embargo, una de las características más importantes de las emociones¹³. Una experiencia visual que representa una pelota como azul, redonda y grande no requiere de un estado mental previo para acceder a la pelota. La propia experiencia visual es la que representa o registra la pelota desde el primer momento. Con las emociones no ocurre lo mismo: para poder representar al perro como peligroso o el gesto de mi hermana como humillante la emoción necesita otros estados mentales para acceder su objeto intencional (por ejemplo, la percepción del perro y el recuerdo del gesto de mi hermana). Esta objeción, sin embargo, sólo estaría mostrando una diferencia relevante entre emociones y percepciones si aceptamos previamente el modelo de la percepción en la que no participan o median otro tipo de estados mentales y en el que el reconocimiento de los objetos de nuestro entorno depende enteramente de percepciones pasivas por parte del sujeto.

La segunda objeción se centra en la fenomenología de percepciones y emociones. El hecho de que tanto emociones como percepciones poseen una importante cualidad sentida está fuera de toda discusión, pero la cuestión

¹² El enactivismo, en el que no nos adentraremos por cuestiones de espacio, ofrece una crítica similar a las teorías clásicas de la percepción. Para una revisión de las teorías enactivistas más importantes en relación a las teorías de las emociones actuales ver Campeggiani (2021).

¹³ En terminología técnica, se dice que las percepciones no tienen “bases cognitivas” mientras que las emociones sí.

relevante, y nada obvia, es si podemos establecer una analogía entre el *tipo* de fenomenología característica de cada uno de estos dos fenómenos. Según Mitchell (2021), la fenomenología de las experiencias perceptivas puede describirse como una fenomenología “presentacional”, en la que algo *se le hace manifiesto* al sujeto: las experiencias perceptivas se sienten como un encuentro con una parte de nuestro entorno que se presenta como siendo de una manera determinada. La fenomenología de las emociones, no obstante, parece ser de otro tipo. Una manera de describirla es en términos de una fenomenología “de respuesta”. La idea fundamental es que las emociones no se sienten como un mero encuentro con una serie de propiedades evaluativas, sino como una *respuesta* a tales propiedades:

La elegancia de sus movimientos me *conmovió*; su triste situación me *afectó*; me sentí *abrumado* por la impresionante vista; su compasivo discurso me *estremeció*; el impactante accidente de coche me *detuvo* (...) hablar de *estar conmovido* por la elegancia de los movimientos de la bailarina, o *afectado* por la grosería de un amigo, sugiere que lo que nos commueve es la gracia de la bailarina o la grosería del amigo, es decir, la manera en la que el objeto particular parece ser evaluativamente. De hecho, es constitutivo de la fenomenología de ser afectado por los valores [o las propiedades evaluativas] el que me muevan en la manera en la que lo hacen¹⁴

Este fragmento parece ofrecer una buena explicación de cómo se sienten nuestras emociones. Podemos preguntarnos, no obstante, si la descripción de la fenomenología de la percepción como meramente “presentacional” es del todo plausible. Según esta distinción, cuando sentimos emociones estamos activamente involucrados en la respuesta emocional mientras que cuando percibimos algo, registramos de manera pasiva los objetos que nos rodean. Cuando percibimos objetos de nuestro entorno, no obstante, no sólo percibimos pelotas azules o grandes, sino que, principalmente, percibimos caras conocidas, gestos cansados, melodías agradables, butacas cómodas, etc. En estos casos, no parece que simplemente entremos en contacto con caras familiares o gestos cansados sino que, más bien, (como en el caso

¹⁴ “the grace of her performance *moved* me; her sorrowful state of affairs *affected* me; I was *overwhelmed* by the stunning vista; his compassionate speech *touched* me; the shocking car crash *arrested* me (...) talk of *being moved* by the gracefulness of the ballerina’s performance, or *affected* by the rudeness of a friend, is suggestive that what we are moved by is the ballerina’s gracefulness or the friend’s rudeness, that is by how the particular object evaluatively seems. Indeed, it is constitutive of *what-it-is-like* to be affected by value to be moved in the way I am” (MITCHELL 2021, énfasis en el original).

de las emociones) la familiaridad de una cara o el cansancio de un gesto nos interpelan o afectan de una manera muy determinada: nos vemos movidos por ellos. Una vez más, por tanto, esta objeción sólo puede convencernos si aceptamos previamente el paradigma de la percepción presupuesto en esta discusión.

Otra de las objeciones planteadas a la teoría perceptiva sostiene que, mientras que las percepciones pueden estar dirigidas a sus objetos (pelota) y a las propiedades de éstos (azul, redonda, grande), las emociones no pueden estar dirigidas a las propiedades de sus objetos. Cuando sentimos miedo ante un perro, nuestro miedo no es una representación de *la peligrosidad*, sino de *el perro* como peligroso. No obstante, la teoría perceptiva defiende que las emociones son percepciones de valor o propiedades evaluativas. Pero si esto fuera así, habríamos de poder decir que, igual que percibimos el azul de la pelota, tememos la peligrosidad del perro. Lo que esta objeción pretende mostrar es que el objeto de las emociones —aquellos a lo que las emociones están dirigidas— no son las propiedades evaluativas sino sus portadores (DEONNA y TERONI 2012). Por tanto, las emociones no pueden ser percepciones de propiedades evaluativas. Al hilo de lo que comentábamos respecto a la objeción anterior, podríamos preguntarnos si, en el caso de percepciones más complejas, cuando percibimos caras conocidas, gestos cansados o butacas cómodas, podemos percibir exclusivamente las propiedades de sus objetos (del mismo modo que podemos percibir el azul de una pelota) o si, como en el caso de las emociones, lo que hacemos es percibir una serie de objetos bajo un determinado aspecto. Está claro que percibimos la familiaridad de una cara o el cansancio de un gesto, pero no parece que podamos concebir esa familiaridad o ese cansancio separado de la propia cara o el propio gesto.

La conclusión es que la teoría perceptiva de las emociones así como las objeciones planteadas en su contra presuponen una determinada concepción de la percepción que condiciona seriamente el tipo de cosas que se pueden decir sobre las emociones. Ante una visión tan controvertida del funcionamiento de la percepción, la manera en la que esta teoría intenta dar cuenta del aspecto intencional, evaluativo y sentido de las emociones no resulta muy convincente.

4.3 Las teorías del sentir

Vamos a adentrarnos ahora en unas teorías cuyo especial énfasis en el aspecto sentido de las emociones ha hecho que se las conozca como las teorías del sentir. En ocasiones, es posible encontrar a algunas de ellas clasificadas de manera distinta, como teorías de la percepción indirecta o como teorías híbridas.

das, pero debido a su estrecha continuidad con la teoría jamesiana, la primera teoría del sentir, consideraremos que todas pertenecen a esta última familia.

4.3.1 *La teoría jamesiana*

La tesis jamesiana es que las emociones son *percepciones de cambios corporales*. Esto quiere decir que, del mismo modo que las teorías perceptivas, las emociones son entendidas como un tipo de percepción. Según James, es un error pensar que la percepción de un hecho (por ejemplo, que un perro me está gruñendo) provoca en nosotros una emoción (miedo) y que esta desencadena una serie de cambios corporales (temblor, sudoración, parálisis, etc.). Los cambios corporales, defiende James, no son efectos causales de las emociones sino que se siguen, más bien, directamente de la percepción del hecho en cuestión (que el perro me está gruñendo). La percepción de esos cambios corporales (el temblor, por ejemplo) *es lo que constituye la emoción*. Esta idea le lleva a afirmar que “nos sentimos tristes porque lloramos, enfadados porque golpeamos, asustados porque temblamos, y no (...) lloramos, golpeamos o temblamos porque, según el caso, estemos tristes, enfadados o asustados (...) En ausencia de los estados corporales que se siguen de la percepción, esto último [la tristeza, el enfado y el miedo] sería puramente cognitivo, pálido, sin color, destituido de toda calidez emocional”¹⁵ (JAMES 1884, 190).

El punto de partida de James es que las emociones *son sentidas*. Según él, hablar de emociones sin hacer referencia a cómo se sienten carece totalmente de sentido: en el momento en que reducimos las emociones a puras cogniciones, las perdemos de vista. Habitualmente, la teoría jamesiana es objeto de duras críticas por las consecuencias que se derivan de invertir, supuestamente, el orden causal del proceso en el que están involucradas nuestras emociones (PINEDA 2019). Si estamos enfadados porque golpeamos, por ejemplo, y no porque alguien nos haya ofendido (lo cual constituiría el hecho desencadenante), no parece que podamos explicar el vínculo entre la emoción y las acciones relacionadas. La emoción del enfado no podría hacer inteligible nuestra acción de confrontar a la persona que nos ha ofendido o nuestra acción de evitarla. Tendríamos que hacer referencia a los cambios corporales producidos por la situación, pero no parece que exista ningún tipo de vínculo explicativo entre golpear y confrontar a una persona.

¹⁵ “we feel sorry because we cry, angry because we strike, afraid because we tremble, and (...) we [don’t] cry, strike, or tremble, because we are sorry, angry, or fearful (...) Without the bodily states following on the perception, the latter would be purely cognitive in form, pale, colourless, destitute of emotional warmth” (JAMES 1884, 190).

Aunque esta sea una objeción importante a la teoría jamesiana, tal vez es posible interpretar la inversión del sentido común que propone James de manera distinta. Podemos, de forma alternativa, considerar que cuando James defiende que “nos sentimos tristes porque lloramos, enfadados porque golpeamos, asustados porque temblamos” el tipo de explicación que está proponiendo no es una explicación de tipo causal sino *constitutiva* (DEONNA y TERONI 2012, 64). Tras rechazar la idea de que las sensaciones corporales son meros efectos causales de las emociones (como hacen las teorías cognitivistas), James trataría de explicar tales sensaciones como *manifestaciones* de las emociones¹⁶. La idea es que el llanto es constitutivo del estar triste: la tristeza que no es sentida, no es tristeza, sino algo puramente cognitivo.

James reclama, en cualquier caso, la centralidad del cuerpo en las emociones. Esto puede considerarse un importante avance en nuestra comprensión del fenómeno, dados los problemas que, como hemos visto, se derivan de dejar el aspecto sentido de las emociones sin explicar o relegado a un segundo plano. La propuesta de James, no obstante, al identificar las emociones con la percepción o sensación de cambios corporales, se enfrenta una serie de dificultades importantes. Una de ellas es que parece que existen sujetos que, debido a una lesión de la médula espinal, no pueden sentir su cuerpo y que, no obstante, disfrutan de una vida emocional muy rica (CANNON 1927; COBOS *et al.* 2002)¹⁷. Según esto, aunque las sensaciones de los cambios corporales son una parte importante de las emociones —o, al menos, de la experiencia típica de las emociones— tales sensaciones no son necesarias para tener emociones.

El problema más grande de la teoría de James es que el tipo de intencionalidad que atribuye a las emociones es claramente erróneo (PINEDA 2019; MITCHELL 2021). Si las emociones son percepciones de cambios corporales, entonces tales cambios constituyen el objeto intencional de las emociones. Decir que el miedo no es otra cosa que la percepción de cómo temblamos, se nos eriza el vello, etc. equivale a decir que el miedo está intencionalmente dirigido a tales cambios. Esta manera de comprender las emociones, sin embargo, es contraria a nuestra manera ordinaria de relacionarnos con ellas. Como decíamos en la sección dedicada a la intencionalidad de las emociones, estas son respuestas a objetos o situaciones que pueden ser evaluadas como respuestas proporcionales o desproporcionales. Dado que, según James, las emociones son respuestas a cambios corporales, plantearse si una emoción

¹⁶ Aunque esta interpretación ayudaría a la teoría jamesiana a esquivar la objeción anterior, esta es una lectura que se aparta de la letra de la propuesta de James (1884). Agradezco a un revisor anónimo esta puntualización.

¹⁷ Para un comentario de esta objeción ver Mitchell (2021, cap. 5).

es una respuesta apropiada o no equivaldría a plantearse si las emociones representan apropiadamente o no dichos cambios corporales, lo cual carece de sentido.

La teoría neo-jamesiana de Prinz (2004), la cual no analizaremos en profundidad por razones de espacio¹⁸, considera que esta última objeción puede evitarse si adoptamos una concepción teleosemántica de la intencionalidad. La idea es que el hecho de que las emociones sean percepciones de cambios corporales no implica que las emociones *representen* o estén dirigidas a tales cambios. Los cambios corporales típicos del miedo (temblor, parálisis) parecen constituir un buen indicador de que estamos en presencia de algún tipo de peligro, del mismo modo que el ascenso del mercurio en un termómetro parece constituir un buen indicador de que la temperatura ha aumentado. Tanto los cambios corporales como el ascenso del mercurio son buenos indicadores de lo anterior porque son causados de manera sistemática por el peligro y por el aumento de temperatura, respectivamente. Cuando la capacidad que una cosa tiene de indicar la presencia de otra cosa ha sido *seleccionada*, adaptativa o artificialmente, por su contribución a nuestras vidas (el temblor como indicador del peligro o el ascenso del mercurio como indicador del aumento de temperatura) podemos decir que tal cosa (el temblor o el mercurio) tiene *la función de llevar información* sobre esa otra cosa (el peligro o la temperatura). Los defensores de la concepción teleosemántica de la intencionalidad sostienen que cuando algo tiene la función de llevar información sobre otra cosa, parece razonable decir que lo primero no indica simplemente la presencia de lo segundo, sino que *lo representa*. Esto quiere decir que el ascenso del mercurio representa una subida de temperatura y que los cambios corporales típicos del miedo representan el peligro¹⁹.

Según esta visión de la intencionalidad, la idea de que las emociones son percepciones de cambios corporales es compatible con la idea de que estas representan propiedades evaluativas. La teoría jamesiana podría, por tanto, reformularse en este sentido y esquivar, en principio, una de sus objeciones más importantes. No obstante, mencionaremos brevemente dos problemas importantes asociados a esta propuesta. En primer lugar, en la forma de representación descrita por Prinz no parece haber lugar para un sujeto que se relaciona intencionalmente, desde una perspectiva de primera persona, con las propiedades evaluativas asociadas a cada tipo de emoción:

¹⁸ Para un análisis detallado de la propuesta de Prinz ver Pineda (2018, cap. 6).

¹⁹ Aunque existen diferentes versiones de la concepción teleosemántica de la intencionalidad, la versión aquí presentada es la más extendida entre los defensores de la teleosemántica (DRETSKE 1988; SHEA 2007; MARTÍNEZ 2019).

Las sensaciones [de miedo] (...) representan el peligro exactamente como el sonido de una alarma (...) representa un allanamiento. Del mismo modo que el sonido de la alarma no está dirigido a quién sea que active la alarma, las sensaciones tampoco están dirigidas al asaltante ni a alguien, como el repartidor, que pueda ser confundido con él. El punto importante es que las sensaciones [corporales] no explican la acción que puedes llevar a cabo en respuesta a la amenaza a la que te enfrentas (...) *no son una evaluación que tú hagas en reacción a ella*²⁰

Como veíamos al principio, sin embargo, este es uno de los presupuestos irrenunciables de nuestro análisis si pretendemos dar cuenta de la experiencia humana y el discurso normativo sobre emociones. En segundo lugar, hay una objeción fenomenológica importante a esta propuesta. La idea es que la conexión entre la fenomenología de las emociones y las propiedades evaluativas correspondientes no parece, tal y como se desprende del análisis de Prinz, en absoluto arbitraria:

No podemos concebir la conexión entre, por ejemplo, la fenomenología del miedo y el peligro como algo arbitrario. Intuitivamente, ninguna otra experiencia emocional que la del miedo es una buena candidata para presentar el mundo en términos de peligro. Aunque no tenemos ningún problema en comprender el hecho de que podemos acabar reaccionando a una situación peligrosa con diversión, digamos, no es inteligible decir que la diversión hace que el peligro de la situación se nos manifieste (...) la conexión entre la experiencia emocional y la propiedad evaluativa no puede modelarse a partir de la conexión que existe entre el humo y el fuego, es decir, como una conexión de co-variació natural. Experimentar la propiedad evaluativa de un objeto no equivale a tomar la manera en la que se siente el cuerpo como una indicación, un signo o un síntoma del hecho de que este objeto tiene esta propiedad²¹

²⁰ “the feelings (...) represent danger exactly as the sound made by an alarm (...) represents a home invasion. Just as the alarm’s sound is not directed at or toward whoever sets off the alarm, so the feelings are not directed at or toward the assailant or anyone, like the deliveryman, mistaken for one. The upshot is that the feelings do not explain whatever action you take in response to the threat you face (...) *they are not an evaluation you make in reaction to it*” (DEIGH 2010, 37, énfasis añadido).

²¹ “we cannot conceive of the connection between, for instance, the phenomenology of fear and danger as arbitrary. Intuitively, no other emotional experience than that of fear is a suitable candidate for presenting the world in terms of a danger. Although we have no problem in understanding the fact that we might end up reacting to a dangerous situation with amusement, say, it is not intelligible to say that amusement makes the danger in the situation manifest to us (...) the connection between the emotional experience and the evaluative property cannot be modeled on that between smoke and fire, namely as one of natural co-variation. Experiencing the evaluative property of an object is not taking the way one’s body feels as an indication, a sign, or a symptom of the fact that this object has this property” (DEONNA y TERONI 2012, 86-7).

En este fragmento se expresa la intuición de que existe una cierta *inteligibilidad primitiva* entre las emociones y las propiedades evaluativas correspondientes. Este tipo de inteligibilidad, no obstante, desaparece por completo si, como sostiene Prinz, decimos que la relación entre el peligro y el temblor es que lo primero causa sistemáticamente lo segundo. No parece, por tanto, que desde una concepción teleosemántica de la intencionalidad se pueda evitar la objeción más importante a la teoría de James.

4.3.2 La teoría actitudinal

A diferencia de las teorías que hemos visto hasta ahora, Deonna y Teroni (2012; 2020) consideran que las emociones no pueden reducirse o explicarse en términos de otro tipo de estado mental. Habitualmente, sostienen ellos, distinguimos entre distintos tipos de estado mental en virtud de que cada uno de ellos constituye *un tipo diferente de actitud* hacia un contenido. La creencia es una actitud que consiste en tomar una proposición o estado de cosas como verdadero; el deseo, una actitud que consiste en tomarlo como algo que debería darse; la suposición, en tomarlo como posible, etc. La idea es que estos diferentes tipos de actitud pueden dirigirse hacia los mismos contenidos: puedo creer que llueve, desear que llueva o suponer que va a llover. Según Deonna y Teroni, las emociones son también diferentes actitudes hacia un contenido: tememos que llueva, nos entristece que llueva, celebramos que llueva, nos frustra que llueva, etc. En lugar de explicar las emociones en términos de otros estados mentales, hemos de reconocer que cada emoción constituye un tipo de actitud diferente hacia un contenido, igual que las creencias, deseos o suposiciones.

Esta teoría sostiene que el tipo de actitud que son las emociones tienen dos características fundamentales. La primera es que se tratan de actitudes *evaluativas*. Cada emoción se corresponde con una actitud evaluativa diferente: el miedo es una actitud que consiste en tomar un objeto o estado de cosas como peligroso, la alegría es una actitud que consiste en tomarlo como un motivo de celebración; el enfado, en tomarlo como algo ofensivo, etc. La segunda característica del tipo de actitud que son las emociones es que se tratan de actitudes *sentidas*. En concreto, las emociones son actitudes *corporales* sentidas. Según Deonna y Teroni, una emoción consiste en un posicionamiento corporal dirigido hacia el objeto de la emoción; un posicionamiento que podemos interpretar en términos de preparación para la acción (FRIJDA 2007). La idea fundamental es que la evaluación del objeto —como peligroso, degradante, ofensivo, etc.— se manifiesta en la propia fenomenología de la emoción, es decir, *en la manera en la que el cuerpo se prepara a actuar* en re-

lación al objeto de la emoción. En términos de la teoría actitudinal, el miedo al perro es una experiencia del perro como peligroso porque consiste en sentir el cuerpo prepararse para disminuir el impacto del perro, el enfado con mi hermana es una experiencia de mi hermana como ofensiva porque consiste en sentir el cuerpo prepararse para tomar represalias contra ella, avergonzarse de una misma es una experiencia de una misma como degradada porque consiste en sentir el cuerpo prepararse para que la tierra le trague o desaparecer de la vista de los demás, etc. En todos estos casos, “sentimos la forma en la que nuestro cuerpo se orienta hacia el objeto en cuestión. El cuerpo se siente en la forma de una gestalt de sensaciones corporales, que consiste en estar preparado a responder de una manera determinada al objeto”²² (2012, 87). Aunque la fenomenología aquí contemplada es de tipo corporal, estos autores rechazan la idea de que podamos explicar cómo se sienten las emociones haciendo referencia exclusivamente a cambios corporales entendidos de forma atomística²³.

Uno de los objetivos de Deonna y Teroni es explicar la manera en la que *las emociones nos ponen en contacto con las propiedades evaluativas* del objeto de

²² “we feel the way our body is geared towards the object we are facing. The body is felt in the form of a gestalt of bodily sensations, which consists in being ready to respond in a given way to the object” (DEONNA y TERONI 2012, 87).

²³ Alguien podría señalar que, al afirmar que las emociones son actitudes sentidas y comportamientos corporales, Deonna y Teroni están haciendo referencia a *dos cosas distintas*, y que los comportamientos corporales son el objeto intencional de las actitudes sentidas. Tal y como entiendo la propuesta de estos autores, no obstante, no parece que se esté hablando de dos elementos separados: las actitudes evaluativas *son* los comportamientos corporales, es decir, al posicionarnos corporalmente en relación al objeto de la emoción *evaluamos* dicho objeto, adoptamos una actitud evaluativa hacia él. “Los cambios corporales emocionalmente relevantes se experimentan como distintos posicionamientos [stances] que adoptamos hacia objetos específicos (...) el sujeto experimenta holísticamente su cuerpo como tomando una actitud hacia un determinado objeto” [the emotionally relevant bodily changes are experienced as distinct stances we adopt towards specific objects (...) the subject experiences her body holistically as taking an attitude towards a certain object] (2012, 79). Respecto a la idea de que los comportamientos corporales podrían ser el objeto intencional de las actitudes evaluativas que constituyen las emociones, Deonna y Teroni rechazan explícitamente esta interpretación e insisten en que tales actitudes corporales sentidas están intencionalmente dirigidas al objeto de la emoción y no a cambios corporales: “la propuesta no es que las emociones sean actitudes que tomamos hacia lo que sentimos que sucede en nuestro cuerpo, sino que lo que sentimos que sucede en nuestro cuerpo constituye en sí mismo una actitud emocional. Las emociones no son actitudes que tomamos hacia nuestro cuerpo sentido, sino actitudes corporales sentidas dirigidas hacia el mundo” [the proposal is not that emotions are attitudes we take towards what we feel happens in our body, but that what we feel happens in our body constitutes in itself an emotional attitude. Emotions are not attitudes we take towards our felt body, but felt bodily attitudes directed towards the world] (2012, 80). No obstante, según Mitchell (2021), estos autores deberían aclarar cómo nos relacionamos con lo que pasa en nuestro cuerpo. Le doy las gracias a un revisor anónimo por animarme a aclarar esto.

la emoción, las cuales, según ellos, sobrevienen de propiedades naturales²⁴. Esta es una cuestión que, desde el punto de vista de estos autores, ni la propuesta de Prinz ni la teoría perceptiva consiguen explicar dadas las objeciones que hemos mencionado antes. Según ellos, los posicionamientos corporales que constituyen las diferentes emociones están intencionalmente dirigidos al objeto de la emoción y en la fenomenología propia de cada tipo de emoción *se manifiestan* las diferentes propiedades evaluativas: sentir cómo mi cuerpo se prepara para huir equivale a experimentar el peligro o sentir cómo mi cuerpo se prepara para que la tierra me trague equivale a experimentar la degradación. Deonna y Teroni añaden que estas actitudes sentidas son actitudes *correctas* únicamente en el caso en que el objeto de la emoción sea de hecho peligroso o degradante. Así, y dada la tesis de la supervenencia, la manera en la que mi cuerpo se prepara para huir del perro será correcta si el perro tiene, por ejemplo, dientes afilados.

Como podemos ver, los aspectos intencional, evaluativo y sentido de las emociones son indisociables según la teoría actitudinal. Los posicionamientos corporales que adoptamos (y que constituyen nuestras emociones) están dirigidos al objeto de la emoción, se relacionan evaluativamente con él (a través de las tendencias específicas a la acción) y son sentidos. Siguiendo a James, esta teoría otorga un papel central al cuerpo en su análisis de las emociones, pero a diferencia de él, el actitudinalismo considera que las emociones no son percepciones de cambios corporales sino actitudes o posicionamientos corporales y sentidos intencionalmente dirigidos al objeto de la emoción. A diferencia de la respuesta de Prinz al problema de la intencionalidad que afectaba a la teoría de James, por otra parte, la teoría actitudinal no toma la manera en la que se siente el cuerpo durante una experiencia emocional como un mero síntoma que indica la presencia de una propiedad evaluativa. Las tendencias a actuar propias de cada emoción son, en principio, inteligibles a la luz de cada tipo de propiedad evaluativa.

La teoría actitudinal, no obstante, es vulnerable a una de las objeciones más importantes a la propuesta de James. Como decíamos, parece que el hecho de no sentir sus cuerpos no constituye un impedimento para que sujetos con una lesión de la médula espinal sientan emociones (CANNON 1927; COBOS *et al.* 2002). Al margen de esta objeción, no podemos dejar de preguntarnos si los posicionamientos corporales que, según estos autores, constituyen las emociones son realmente inteligibles a la luz de las diferentes propiedades evaluativas. En concreto, si estos diferentes posicionamientos pueden ser evaluados como correctos o incorrectos en relación a las distintas propiedades:

²⁴ Según la tesis de la supervenencia, propiedades evaluativas como la peligrosidad de un perro, por ejemplo, sobrevienen de propiedades naturales como tener dientes afilados, una gran fuerza física, estar en posición de ataque, etc. (KIM 1993).

Ante la pregunta de a qué equivale la corrección en la explicación que Deonna y Teroni dan del miedo, es natural responder “racional”. Plausiblemente, es racional huir si uno detecta una amenaza e irracional si este no es el caso. Pero esta respuesta es mucho menos plausible en los demás casos. ¿Es racional prepararse para tomar represalias cuando nos ofenden? No es ni mucho menos obvio que lo sea. ¿Es racional desaparecer en el suelo cuando se nos degrada? Estoy tentado de decir que, en la medida en que puedo encontrarle sentido a esta sugerencia, no lo es. Sin embargo, más importante que esta perplejidad es la sospecha de que la corrección en estos análisis es, en última instancia, una cuestión de prudencia o de propiedad moral. Pero estas nociones, tal y como presuponen Deonna y Teroni, deben contrastarse con la corrección²⁵

El problema es que resulta muy difícil describir patrones de posicionamientos corporales que se correspondan de manera unívoca con las diferentes propiedades evaluativas. Además, el peligro, más que merecer que nuestro cuerpo se prepare para huir, lo que merece es miedo. La pregunta es si el miedo puede ser reducido a cómo nuestro cuerpo se prepara para huir. Mi cuerpo podría, por ejemplo, prepararse para huir ante una amenaza y yo no experimentar ningún tipo de miedo, y alguien podría decirme que, si no siento miedo, es que hay algún aspecto de la situación que no estoy captando bien.

Otra perplejidad de esta teoría es la idea de que tener una emoción es sentir cómo *el cuerpo* se prepara para responder. Como hemos insistido en diversas ocasiones, los que respondemos emocionalmente a algo, a partir de nuestra perspectiva evaluativa sobre la situación en la que nos encontramos, *somos nosotros*. Si no, no tendría sentido que nos corrigiéramos en la manera en la que lo hacemos (“no deberías sentirte culpable por esto”, “el comentario que le hizo fue muy ofensivo, normal que se enfadara”, “no entiendo cómo no derramó ni una lágrima durante el funeral”, etc.). Tal vez, podríamos decir que sentir miedo es sentir cómo *nosotros* nos preparamos para huir, y no nuestro cuerpo. No obstante, las prácticas que regulan nuestras emociones no parecen en absoluto estar regulando meramente posicionamientos corporales.

²⁵ “To the question of what correctness amounts to in Deonna and Teroni’s account of fear, it is natural to answer “rational”. Indeed, plausibly, it is rational to flee if one spies a threat and irrational otherwise. But such an answer is markedly less plausible in the other cases. Is it rational to prepare for retaliation when offended? It is far from obvious that it is. Is it rational to disappear into the ground when degraded? I am tempted to say that, insofar as I can make sense of the suggestion, it is not. More significant than such puzzlement, however, is the suspicion that correctness in these analyses is ultimately a matter of prudential or moral appropriateness. But these notions, so Deonna and Teroni suppose, are to be contrasted with correctness” (SMITH 2014, 11).

Varios teóricos de las emociones se han alejado de este tipo de propuesta argumentando que las emociones —en concreto, su cualidad sentida— no pueden reducirse a cómo se siente el cuerpo (GOLDIE 2000, HELM 2002; MITCHELL 2021). Según estos autores, tenemos que hacer referencia a otro tipo de sentires o sensaciones que no sean meramente corporales para entender la manera en la que las emociones se dirigen a su objeto y nos ponen en contacto con sus propiedades evaluativas. La idea es que no podemos separar entre el sujeto y su cuerpo cuando nos aproximamos a sus emociones: “toda nuestra mente y nuestro cuerpo están involucrados en la experiencia emocional, y todos los sentires están ‘unidos en la conciencia’ en su estar dirigidos hacia el objeto: unidos ‘en cuerpo y alma’”²⁶ (GOLDIE 2000, 55). El problema es que los intentos por desarrollar este nuevo proyecto²⁷, del mismo modo que la teoría actitudinal, dan una explicación poco convincente de cómo este tipo de actitud especial —que no es meramente corporal— nos da acceso a las propiedades evaluativas del objeto de la emoción. La propuesta más habitual es que nuestras emociones están siempre asociadas a una determinada valencia: al experimentarlas, parte de lo que sentimos es aprobación o desaprobación por el objeto de la emoción (MITCHELL 2021) o sensaciones de placer o dolor en relación a dicho objeto (HELM 2002).

Tomemos el caso del enfado: la experiencia incluye una respuesta de *desaprobación sentida*, que representa afectivamente el desvalor [*disvalue*] del comentario ofensivo. O tomemos el caso de la diversión: la experiencia incluye una respuesta de *favor sentida*, que representa afectivamente el valor de la broma divertida. Por último, tomemos el caso de la admiración: la experiencia incluye una respuesta de *favor sentida*, que representa afectivamente el valor de la actuación elegante. Así pues, tenemos un candidato a ser el componente que puede captar de una vez el sentido en que la experiencia emocional incluye el *ser movido por el valor*, y al mismo tiempo representa afectivamente ese mismo valor²⁸

²⁶ “our entire mind and body is engaged in the emotional experience, and all the feelings are ‘united in consciousness’ in being directed towards its object: united ‘body and soul’, ‘heart and mind’” (GOLDIE 2000, 55).

²⁷ El cual ha recibido recientemente el nombre de teorías de la intencionalidad afectiva (VENDRELL 2018).

²⁸ “Take the case of anger: the experience includes a response of *felt disfavour*, which affectively represents the disvalue of the offensive remark. Or take the case of amusement: the experience includes a response of *felt favour*, which affectively represents the value of the funny joke. Finally, take the case of admiration: the experience includes a response of *felt favour*, which affectively represents the value of the graceful performance. As such, we have a candidate for a component which claims to in one stroke capture the sense in which emotional experience includes being *moved by value*, and at the same time affectively represents that very value” (MITCHELL 2021, énfasis en el original).

Como vemos, la complejidad de la cuestión no hace más que aumentar. La pregunta que puede surgirnos a leer este fragmento es cómo es posible que sensaciones afectivas tan básicas como aprobar o desaprobar nos pongan en contacto con el conjunto de propiedades evaluativas que se corresponden con los diferentes tipos de emoción. Este tipo de actitudes podrían, en todo caso, ponernos en contacto con propiedades más simples o básicas (como tener valor o no tenerlo) pero no parece que la experiencia del peligro, por ejemplo, pueda ser explicada exclusivamente en términos de una desaprobación sentida. Cuando siento miedo ante un perro, siento que es peligroso, no que carece de valor para mí.

5. EL PROBLEMA DE LA RELACIÓN ENTRE EMOCIONES Y PROPIEDADES EVALUATIVAS

Hasta aquí llega nuestra revisión de las diferentes teorías de las emociones. De la discusión llevada a cabo se desprende que la preocupación más importante de los participantes en este debate es esclarecer *el tipo de estado mental* que son las emociones. Sin embargo, tal y como hemos podido ver, la identificación de las emociones con creencias y deseos, juicios evaluativos, percepciones (de propiedades evaluativas o cambios corporales) o actitudes evaluativas sentidas resulta problemática por los motivos que hemos comentado a lo largo de la sección anterior.

Parece haber, por otra parte, un problema más profundo y complejo que los tratados hasta ahora y que condiciona seriamente el debate actual sobre emociones. Se trata de una cuestión que afecta especialmente a la teoría actitudinal pero que determina asimismo las propuestas de la teoría preceptiva y de las teorías de la intencionalidad afectiva a las que nos hemos referido brevemente en la sección anterior. Como hemos visto, una de las preocupaciones más importantes de teóricos como Deonna y Teroni, Tappolet o Mitchell es esclarecer la manera en la que las emociones nos permiten acceder a las propiedades evaluativas de los objetos a los que están dirigidas. Esta preocupación centra la investigación de estos autores en la búsqueda de un estado mental —*sui generis* o no— que, entre otras características, tenga, precisamente, *la capacidad de registrar o dar acceso a las propiedades evaluativas* del objeto de la emoción. Tal y como expresaban Deonna y Teroni, intuitivamente, parece que hay algo así como una inteligibilidad primitiva entre las emociones y sus correspondientes propiedades evaluativas —“ninguna otra experiencia emocional que la del miedo es una buena candidata para presentar el mundo en términos de peligro” (2012, 86)— por lo que, según ellos, una buena teoría

no puede tomar, como hace Prinz, los cambios corporales como meros síntomas de la presencia de una propiedad. El objetivo es, por tanto, explicar cómo las propiedades evaluativas se nos manifiestan en experiencias emocionales de primera persona.

La idea de que hemos de encontrar un tipo de estado mental que pueda proporcionarnos acceso a propiedades evaluativas y que podamos identificar con las emociones presupone, no obstante, una determinada concepción de las *propiedades evaluativas como independientes de nuestras respuestas emocionales*. A continuación, revisamos brevemente este compromiso filosófico y nos planteamos si las dificultades a las que se enfrentan varias de las teorías que hemos examinado aquí tienen más que ver con la adopción de este compromiso que con la supuesta complejidad de las emociones.

5.1 Las propiedades evaluativas como independientes de las emociones

Los filósofos que piensan que las propiedades evaluativas son independientes de nuestras respuestas emocionales consideran que las propiedades evaluativas son propiedades del mundo —que se encuentran ahí fuera— que nuestras emociones registran y que supervienen de propiedades naturales (DEONNA y TERONI 2012; TAPPOLET 2016; MITCHELL 2021). Así, la propiedad de ser peligroso de un perro sobrepondría de la propiedad natural de tener colmillos afilados, un gruñido fuerte, una gran fuerza física, etc. Si alguna de estas propiedades cambiara, el perro dejaría de ser peligroso. Igualmente, que tu entorno te ignore, no te mire a los ojos o te llame por un nombre que no es el tuyo son todas propiedades naturales (en un sentido muy laxo) de las que sobreviene la propiedad de ser humillante. Cuando nos preguntamos si algo es peligroso o humillante, lo que hemos de hacer, según este planteamiento, es *investigar el mundo y sus propiedades naturales*: si el perro no tiene dientes afilados ni una gran fuerza física, podemos concluir que no es peligroso; si el entorno de una persona no le está ignorando ni llamando por un nombre que no es el suyo, podemos concluir que la persona en cuestión no está recibiendo un trato humillante.

Deonna y Teroni, por ejemplo, sostienen que mi miedo al perro (el cual representa al perro como peligroso) está justificado si mi emoción está basada en la percepción²⁹ de las propiedades naturales de las que sobreviene la peligrosidad del perro. Es decir, mi miedo está justificado si he visto un perro con colmillos grandes. Desde el punto de vista de Smith, no obstante, esta idea resulta problemática porque

²⁹ En la percepción *verídica*.

la justificación se transmite a través del contenido, por lo que el contenido *no evaluativo* de la base cognitiva [el contenido de la percepción visual del perro] (el perro tiene dientes grandes) no puede justificar la actitud *evaluativa* que es la emoción (el perro es peligroso) (...) una representación de algo como F [como de dientes grandes] no puede justificar que uno lo tome como G [como peligroso] en ausencia de alguna información adicional (digamos que todo lo F es G) (...) deberíamos ser muy escépticos con la idea de que ser consciente de que algo tiene dientes grandes, incluso en un contexto en el que esta propiedad constituye una amenaza, sea suficiente para justificar que uno lo considere una amenaza³⁰

Según esta crítica, necesitamos una explicación del salto que parece haber entre lo natural y lo evaluativo: lo que decimos en términos naturales no parece equivaler al tipo de cosas que expresamos en los términos evaluativos sin los cuales es imposible diferenciar los diferentes tipos de emoción.

Como veíamos en el apartado sobre el aspecto evaluativo de las emociones, el miedo no es una respuesta a lo que tiene la propiedad de morder, sino a lo que tiene la propiedad de ser peligroso. Del mismo modo, la humillación no es una respuesta a quien no te llama por tu nombre, sino a quien te humilla. Podemos ir añadiendo propiedades naturales para intentar explicar por qué algo es peligroso o humillante pero parece que cualquier tipo de explicación que conecte las propiedades naturales con las evaluativas incluirá en sí misma términos y conceptos de tipo evaluativo (BERLIN 1958; CORBÍ 2012). Si intentamos explicar por qué la propiedad de ser peligroso sobreviene de la propiedad de morder es inevitable hacer referencia al hecho de que recibir el mordisco de un perro pone en riesgo nuestra integridad física. Pero, entonces, no hemos explicado la propiedad de ser peligroso en términos naturales sino en términos evaluativos, esto es, mediante conceptos como “poner en riesgo” o “integridad”. Lo mismo ocurre con la propiedad de ser humillante. Parece que no podemos explicar por qué el que alguien te ignore, no te mire a los ojos o te llame por un nombre que no es el tuyo constituye algo humillante a menos que contemos una historia sobre ello. El problema, según los defensores de este planteamiento, es que esta historia estará repleta de términos evaluativos, no naturales: que te llamen por un nombre que no es el tuyo es una falta de respeto, que no te miren a los ojos es negarse a reconocer que eres una persona al

³⁰ “justification travels via content, so the *non-evaluative* content of the cognitive base (that the dog has big teeth) cannot justify the *evaluative* attitude that is the emotion (that it is dangerous). According to this principle, a representation of something as an F cannot not justify one’s taking it to be a G in the absence of some further information (say that all Fs are Gs) (...) we should be highly sceptical of the idea that being aware of something as having big teeth, even in a context in which this property constitutes being a threat, is sufficient to justify one’s taking it to be a threat” (SMITH 2014, 104-5, énfasis añadido).

mismo nivel que los demás, etc. Hacer referencia exclusivamente a propiedades naturales, además, tampoco parece explicar por qué algo es humillante en lugar de, por ejemplo, envidiable.

5.2 La relación de dependencia mutua entre emociones y propiedades evaluativas

A diferencia de esta última postura filosófica, hay quien piensa que, del mismo modo que hemos de hacer referencia a las diferentes propiedades evaluativas para individuar los distintos tipos de emoción, no podemos *individuar las propias propiedades evaluativas* sin hacer referencia a los distintos tipos de emoción: lo ofensivo es lo que merece enfado, lo peligroso es lo que merece miedo, lo vergonzoso es lo que merece vergüenza, etc. —y no lo que sobreviene de ciertas propiedades naturales—.

La diversión, por ejemplo, es una reacción que tenemos que caracterizar por referencia a su objeto correspondiente, a través de algo percibido como divertido (...) No hay una explicación de la diversión que sea independiente del objeto, independiente de la propiedad, “puramente fenomenológica” o “puramente introspectiva”. Igualmente, tampoco puede decirse qué es exactamente lo divertido sin hacer referencia a la risa, a la diversión o a reacciones similares. ¿Por qué habríamos de esperar que existiera tal explicación independiente? Por supuesto, cuando discutimos si x es realmente divertido, hay todo un conjunto de consideraciones y explicaciones que podemos aducir, y de ninguna manera todas ellas tienen que ser dadas en términos simplemente sinónimos o interdefinibles con “divertido” (...) Lo que es extremadamente improbable es que (...) explicaciones ulteriores resulten ser una *reducción* de lo divertido o sirvan para caracterizarlo en términos puramente naturales³¹

Este fragmento presupone que existe una relación de *dependencia mutua* entre emociones y propiedades evaluativas. Desde esta perspectiva, la

³¹ “Amusement for instance is a reaction we have to characterize by reference to its proper object, via something perceived as funny (...) There is no object-independent and property-independent, ‘purely phenomenological’ or ‘purely introspective’ account of amusement. And equally there is no saying what exactly the funny is without reference to laughter or amusement or kindred reactions. Why should we expect there to be such an independent account? Of course, when we dispute whether x is really funny, there is a whole wealth of considerations and explanations we can adduce, and by no means all of them have to be given in terms simply synonymous or interdefinable with ‘funny’ (...) What is improbable in the extreme is that (...) further explanations will ever add up to a *reduction* of the funny or serve to characterize it in purely natural terms” (WIGGINS 1987, 195, énfasis en el original).

estrecha conexión entre respuestas emocionales y propiedades – su inteligibilidad primitiva – no es ningún misterio: estos dos elementos son como las dos caras de una misma moneda. A pesar de que diferentes versiones de esta teoría han sido respaldadas por un importante número de filósofos (WIGGINS 1987; GOLDIE 2000; DUNN 2006; CORBÍ 2012), se trata de una teoría que ha sido ampliamente rechazada en el debate actual sobre emociones. La tendencia general, como veíamos antes, es a considerar que las propiedades evaluativas son independientes de las emociones y que la inteligibilidad primitiva entre estos dos elementos es una cuestión que *requiere ser explicada*. Este es el motivo por el que teorías como la actitudinal se centran en explicar la manera especial en la que las emociones registran o nos dan acceso a las propiedades evaluativas.

Los argumentos que se han presentado en contra de la idea de que existe una relación de dependencia mutua entre emociones y propiedades evaluativas, no obstante, no parecen ser definitivos o, al menos, no parecen determinar que una concepción de las propiedades evaluativas como independientes de las emociones sea claramente preferible. Además, como acabamos de ver, hay quien piensa que esta última concepción no está libre de problemas. Dado que la adopción de una determinada concepción de la relación entre emociones y propiedades evaluativas condiciona significativamente la manera en la que nos aproximamos a las emociones, resulta pertinente que revisemos brevemente los argumentos que se han presentado para rechazar la concepción de la dependencia mutua entre emociones y propiedades así como algunas de las respuestas ofrecidas³².

La primera objeción a la idea de que las propiedades evaluativas dependen de las respuestas emocionales para ser individuadas³³ es que esta es una explicación circular. Si decimos que hemos de individuar los distintos tipos de emoción por referencia a las propiedades evaluativas y viceversa, no estamos aportando ningún tipo de información nueva: algo tiene la propiedad de ser encomiable si merece orgullo, y algo merece orgullo si tiene la propiedad de ser encomiable.

³² Para una visión mucho más detallada de los argumentos en contra de la concepción de la dependencia mutua entre emociones y propiedades evaluativas ver Deonna y Teroni (2012, cap. 4) y Pineda (2019, cap. 8).

³³ El hecho de que las propiedades evaluativas hayan de ser individuadas por referencia a nuestras respuestas emocionales no implica que tales propiedades no sean más que meras proyecciones nuestras en lugar de hechos evaluativos del mundo. En otras palabras, esta visión no implica que las propiedades evaluativas sean propiedades dependientes de respuesta. El proyectivismo moral (KORSGAARD 1996) es incompatible con la individuación del contenido evaluativo de nuestras emociones y nuestros juicios evaluativos (CORBÍ 2012, cap. 4). La idea de que existe una relación de dependencia mutua entre emociones y propiedades evaluativas, por tanto, es compatible con el realismo moral.

En respuesta a esta objeción, Wiggins (1987) indica que la circularidad involucrada en esta teoría es *virtuosa*, pues lo que se pretende ofrecer no es una definición o análisis de las propiedades evaluativas sino una elucidación. Es decir, el objetivo de este tipo de explicación no es aportar información nueva, sino poner las propiedades evaluativas en el contexto en el que podemos ver su sentido. La idea principal es que las respuestas emocionales y las propiedades evaluativas han de ser observadas las unas a la luz de las otras o, usando la terminología de Finkeslein (2003), que el par respuesta-propiedad forma una *unidad de inteligibilidad*. Por supuesto, podemos sacar los elementos de este par del contexto al que pertenecen y observarlos por separado. No obstante, si esta se convierte en la manera habitual en la que nos aproximamos a las emociones y a las propiedades evaluativas, la estrecha conexión que existe entre ambas se vuelve un misterio. En el contexto apropiado, en cambio, podemos ver fácilmente cómo las preguntas “¿he de admirar x?” y “¿es x admirable?” son indistinguibles (DUNN 2006).

La segunda objeción a esta postura es que, si las propiedades evaluativas han de ser individuadas por las respuestas emocionales que producen, no puedo estar nunca equivocada con respecto a si algo tiene una determinada propiedad, pues si he sentido la emoción correspondiente, estoy obligada a concluir que de hecho tiene tal propiedad. Por ejemplo, si algo me hace reír, he de concluir que se trata de algo divertido. Nuestro discurso normativo sobre emociones, sin embargo, sugiere que las emociones y las propiedades evaluativas pueden darse por separado: a veces, nos reímos de algo que no es divertido y somos incapaces de llorar ante algo que es profundamente triste.

En el fragmento anterior, veíamos cómo Wiggins afirma que “por supuesto, cuando discutimos si x es realmente divertido, hay todo un conjunto de consideraciones y explicaciones que podemos aducir” (1987, 195). La objeción funcionaría si dijéramos que las propiedades evaluativas se individúan por referencia a nuestras respuestas emocionales *actuales*, pero esta no es la tesis de Wiggins. Según él, a través de un proceso de instrucción mutua y mejoramiento, aprendemos que los objetos y situaciones a los que respondemos con una determinada emoción poseen paradigmáticamente una determinada propiedad evaluativa: aquello a lo que respondemos con miedo es paradigmáticamente peligroso, mientras que aquello a lo que respondemos con vergüenza es paradigmáticamente vergonzoso. Una vez hemos aprendido esto, “otra cosa también se vuelve posible. En lugar de fijarnos en un objeto o en una clase de objetos y discutir sobre qué respuesta o respuestas son las que evocan, podemos fijarnos en una respuesta (...) y luego discutir sobre cuáles son las marcas de la propiedad para las que la respuesta misma

está hecha”³⁴ (WIGGINS 1987, 198). Es decir, una vez alcanzado este estadio, podemos comprobar si algo tiene una determinada propiedad fijándonos en las marcas propias de los casos paradigmáticos y no meramente en la presencia o ausencia de una determinada respuesta emocional. Esto nos permite ver, por una parte, que algo puede ser peligroso aunque no sintamos miedo y, por otra, que algo puede no serlo a pesar de que sí que sintamos miedo³⁵.

La tercera objeción a la idea de que las propiedades evaluativas dependen de las respuestas emocionales es que esta postura conduce irremediablemente al relativismo moral. Si nos preguntamos si cierto acto es humillante, por ejemplo, una persona que se sienta humillada por él concluirá que sí lo es mientras que una persona que no se sienta de tal modo alcanzará la conclusión contraria, *y ambas respuestas serán igualmente correctas*. La consecuencia más importante es que esto nos deja sin la posibilidad del desacuerdo moral, por lo que nuestros debates sobre cuestiones de este tipo carecerían de sentido. No obstante, debatimos constantemente sobre cuestiones morales e intentamos justificar cuál es la respuesta correcta ante una situación determinada.

En respuesta a esta objeción, filósofos como Berlin (1958) o Corbí (2012) insisten en la importancia de diferenciar claramente el relativismo moral³⁶ del *pluralismo moral*. Asumir que ante un determinado problema moral no existe una única respuesta válida, no implica que *cualquier respuesta lo sea*. Cuando deliberamos sobre qué hacer o qué pensar, no nos convertimos en seres abstractos y objetivos que deliberan, sino que partimos de nuestra situación particular, la cual involucra nuestro carácter, preferencias, proyecto vital, historia personal, etc. No es de extrañar, por tanto, que diferentes individuos otorguen un peso diferente a las razones que cuentan a favor o en contra de una opción concreta (WILLIAMS 1976). Esto no quiere decir que cualquier respuesta arbitraria sea válida: para que una respuesta cuente como propiamente moral ha de estar vinculada a aspectos moralmente relevantes de la situación en cuestión. Como señala Corbí, hasta “la persona más racista debe asociar sus ideas sobre la raza con algunas características moralmente relevantes como la inteligencia o la sociabilidad, si queremos interpretar que sus afirmaciones expresan una opinión con la que podría-

³⁴ “Then something else also becomes possible. Instead of fixing on an object or class of objects and arguing about what response or responses they are such as to evoke, we can fix on the response (...) and then argue about what the marks are of the property that the response itself is made for” (WIGGINS 1987, 198).

³⁵ Ver también Pettit (1991).

³⁶ Cuando se habla de relativismo en esta discusión se hace referencia a un tipo de relativismo que se caracteriza por aceptar toda respuesta como válida y no se contemplan relativismos más sofisticados como puede ser un relativismo de carácter local.

mos estar moralmente en desacuerdo”³⁷ (2012, 125). Asumir, por tanto, que existen diferentes respuestas válidas a una pregunta no implica el relativismo ni impide la discusión sobre cuestiones morales. Corbí afirma, además, que la objeción planteada presupone que cuando debatimos sobre este tipo de cuestiones nuestro objetivo no es otro que encontrar cuál es la *opinión verdadera* en cada caso. No obstante, cuando reflexionamos sobre el daño y sobre cómo prevenirlo o repararlo, la cuestión importante es, más bien, *articular una respuesta proporcionada al daño*, y no si nosotros o las personas involucradas sabemos qué es lo correcto. La articulación de una respuesta de este tipo involucra, además, una serie de elementos epistémicos que, según él, no podemos comprender si perdemos de vista el contraste entre relativismo y pluralismo moral.

La cuarta y última objeción a la idea de que las propiedades evaluativas dependen de las respuestas emocionales afirma que, a menos que esta postura aclare el tipo de razones que determinan qué emoción es la que una situación merece, dicha propuesta no es sostenible. Cuando evaluamos si una determinada emoción es adecuada, empleamos diferentes tipos de estándar. En primer lugar, los estándares de corrección nos permiten evaluar las emociones en función de la presencia o ausencia de la propiedad evaluativa correspondiente: el miedo al perro no es correcto si el perro no es peligroso (DEONNA y TERONI 2012). En segundo lugar, los estándares de propiedad establecen si sentir una determinada emoción es moralmente apropiado: aunque una broma sea divertida y, por tanto, sentir diversión sea correcto (según los estándares de corrección) sentir diversión resulta moralmente inapropiado si podemos herir los sentimientos de otra persona al reírnos (D'ARMS y JACOBSON 2000). En tercer lugar, hay estándares de tipo prudencial: aunque admirar a un vil demonio sea incorrecto, dado que no se trata de un individuo admirable, en términos prudenciales conviene que le admiramos si no hacerlo nos pone en algún tipo de riesgo (RABINOWICZ y RØNNOW-RASMUSSEN 2004). Dados estos tres tipos de estándar, podemos concluir que el miedo no es la respuesta que merece el perro, la diversión no es la respuesta que merece la broma y la admiración es la respuesta que merece el demonio. Ahora bien, si las propiedades evaluativas han de ser individuadas por las respuestas emocionales que merecen, en el caso de la broma, dado que esta no merece diversión (según el criterio de propiedad), habríamos de concluir que esta no es divertida. En el caso del demonio, dado que merece admiración (según el criterio de tipo prudencial), habríamos de concluir que se trata de alguien admirable.

³⁷ “the wildest racist must associate her claims about race with some morally relevant features like intelligence or sociability, if we are to interpret her claims as expressing a view we might morally disagree with” (CORBÍ 2012, 125).

Aunque esta objeción suele plantearse como un problema muy serio para los defensores de la relación de dependencia mutua entre emociones y propiedades evaluativas, podemos planteárnos si el análisis de los ejemplos anteriores nos resulta convincente. Por una parte, decir que, a pesar de que el vil demonio no sea un individuo admirable, conviene admirarle por motivos prudenciales, parece un poco forzado. *Mostrar* admiración o evitar mostrar rechazo ante él puede resultar, en efecto, conveniente, pero ¿en qué sentido es útil que *sintamos* admiración por él? Admirarle podría ser, incluso, contrario a nuestros intereses, pues implicaría abandonar nuestras convicciones más profundas (si es que esto es algo que podamos hacer y, más aún, por razones prudenciales). Alguien podría insistir, sin embargo, en que, por motivos de supervivencia, resulta mucho más conveniente sentir admiración por el demonio que fingir simplemente tal emoción, pues en el primer caso estaremos motivados a actuar de manera prudencial³⁸. La cuestión importante es si, dado este caso, el defensor de la dependencia mutua entre emociones y propiedades evaluativas está obligado a concluir que el demonio *es admirable*. Tal y como indicábamos al hilo de la segunda objeción, el defensor de esta postura puede reconocer situaciones en las que un sujeto particular responde con una determinada emoción (admiración en nuestro ejemplo) sin que el objeto al que se dirige su emoción merezca tal emoción. Es decir, el hecho de que un sujeto particular sienta admiración ante un determinado objeto no quiere decir que tal objeto sea admirable³⁹. Podríamos comprender por qué el sujeto en cuestión siente admiración —e incluso pensar que, de encontrarnos en la misma situación, es probable es que sintiéramos lo mismo que él o ella— pero ver, al mismo tiempo, esta forma de admiración como una *desviación* del caso paradigmático. Por lo tanto, no parece que el defensor de la dependencia mutua entre emociones y propiedades esté obligado a concluir que el demonio sea admirable. En el caso de la broma, por otra parte, si todo lo que importara fuera no herir los sentimientos de otra persona, este tipo de consideración, más que de tipo moral, parece ser de tipo prudencial: aunque sentir diversión sea correcto, no debería reírme delante de la persona a quién la broma va dirigida. No obstante, podemos preguntarnos al hilo de la respuesta a la objeción anterior, si esta es la cuestión que realmente importa. Podría ser que el motivo por el que resulta moralmente inapropiado vivir la broma como algo divertido es que ello me insensibiliza ante el daño que esta puede estar causando al otro. Esta idea podría hacernos cuestionar en qué

³⁸ Le agradezco a un revisor anónimo la insistencia en este punto.

³⁹ Aunque tengamos razones prudenciales para admirar a alguien, estas no son el único tipo de razones que cuentan a favor de admirar a alguien. Para determinar si alguien es admirable, es importante considerar *en cada caso* el *conjunto* de razones que tenemos a favor y en contra de admirar al individuo en cuestión.

sentido relevante, entonces, hemos de seguir insistiendo en que la diversión es correcta aunque moralmente inapropiada⁴⁰.

La revisión de estos argumentos, aunque breve, resulta pertinente porque nos hace ver que la relación entre emociones y propiedades evaluativas es una cuestión abierta y que no es evidente que hayamos de rechazar la concepción de la dependencia mutua entre emociones y propiedades en favor de la postura contraria. Sin embargo, y como hemos visto, las teorías de las emociones más recientes están guiadas por la idea de que las propiedades evaluativas son independientes de las emociones y que parte de lo que tenemos que investigar —cuando nos preguntamos qué son las emociones— es de qué manera las emociones registran o dan acceso a tales propiedades. Si recordamos, uno de los problemas más importantes de varias de las teorías que hemos revisado es, precisamente, que su explicación de la manera en la que las emociones nos ponen en contacto con las propiedades evaluativas no resulta especialmente convincente: respecto a la teoría perceptiva, veíamos que decir que percibimos la propiedad de ser peligroso de la misma manera que percibimos la propiedad de ser azul no puede ser correcto; respecto a la teoría actitudinal, no estaba del todo clara la relación entre los posicionamientos corporales descritos y las correspondientes propiedades evaluativas; respecto a las teorías de la intencionalidad afectiva, indicábamos que no parece que sensaciones afectivas tan básicas como aprobar o desaprobar puedan ponernos en contacto con propiedades complejas como ser peligroso o reprochable.

Da la impresión de que todas estas teorías se atascan en su explicación de la conexión entre emociones y propiedades evaluativas. Llegados a este punto, no podemos dejar de preguntarnos si esta dificultad es inherente a la complejidad del fenómeno estudiado o si podría estar relacionada, más bien, con la concepción de las propiedades evaluativas presupuesta por tales teorías. Para un defensor de la dependencia mutua entre emociones y propiedades, por ejemplo, no tendría sentido preguntarse por la manera en la que las emociones nos ponen en contacto con un conjunto de propiedades totalmente independientes de ellas. Probablemente, una teoría de las emociones guiada por esta última concepción trataría de dar respuesta a otro tipo de preguntas. Más allá de optar por una determinada concepción de la relación entre emociones y propiedades evaluativas, el objetivo de esta última sección ha sido mostrar que la adopción de una u otra concepción es una cuestión abierta y de gran relevancia para el debate sobre qué son las emociones. Un aprendizaje importante

⁴⁰ En este punto conviene recordar que el requisito de distinguir entre diferentes tipos de estándar (de corrección, propiedad y de tipo prudencial) que plantean, entre otros, Deonna y Teroni, es un requisito que la propia teoría actitudinal no satisface. Ver la cita de Smith (2014, 11) en la subsección 4.3.2.

es que, cuando examinamos diferentes teorías de las emociones, es importante investigar la manera en la que compromisos filosóficos de este tipo pueden estar condicionando la manera de entender el fenómeno en cuestión.

6. CONCLUSIÓN

Hemos comenzado precisando el tipo de explicación personal que iba a guiar nuestro análisis así como el tipo de fenómeno afectivo en el que se centran los teóricos de las emociones actuales. Tras haber descrito los tres aspectos distintivos de las emociones —intencional, evaluativo y sentido—, hemos revisado las teorías de las emociones más importantes de las últimas décadas. A lo largo de esta revisión, hemos ido identificando una serie de concepciones y presupuestos que parecen condicionar tanto las diferentes teorías como las objeciones planteadas a las mismas. Por último, hemos profundizado en uno de los compromisos filosóficos que más determinan el debate actual sobre emociones, a saber: la concepción de las propiedades evaluativas como independientes de las respuestas emocionales.

Agradecimientos

Parte de las ideas aquí expuestas fueron presentadas por primera vez en el Seminari de Filosofia de la Càtedra Blasco, donde recibí comentarios muy útiles que me ayudaron a desarrollar este artículo. Quiero dar las gracias, en especial, a Josep Corbí y a Jordi Valor por las conversaciones que hemos mantenido sobre diferentes versiones de este texto y que tanto me han ayudado a mejorarlo. Les agradezco también a los editores de esta revista su cuidadosa manera de trabajar y a un revisor anónimo sus detallados comentarios y sugerencias. Para la realización de este artículo he contado con el apoyo del proyecto “Autoconocimiento, responsabilidad moral y autenticidad” del Ministerio de Ciencia, Innovación y Universidades (FFI2016-75323-P). Esta publicación es parte de la ayuda BES-2017-081537, financiada por MCIN/AEI/10.13039/501100011033 y por FSE invierte en tu futuro.

REFERENCIAS BIBLIOGRÁFICAS

- BRADY, M. S. 2009, "The Irrationality of Recalcitrant Emotions", *Philosophical Studies*, 145 (3): 413-30.
- CAMPEGGIANI, P. 2021, *Introduzione alla filosofia delle emozioni*, Biblioteca Clueb.
- CANNON, W. B. 1927, "The James-Lang Theory of Emotion: A Critical Examination and an Alternative Theory", *American Journal of Psychology*, 39: 106-24.
- CARMAN, M. 2018, "How Emotions Do Not Provide Reasons for Action", *Philosophia*, 46: 555-74.
- COBOS, P. et al. 2002, "Revisiting the James versus Cannon Debate on Emotion: Startle and Autonomic Modulation in Patients with Spinal Cord Injuries", *Biological Psychology*, 61 (3): 251-69.
- CORBÍ, J. E. 2012, *Morality, Self-Knowledge and Human Suffering*, Routledge.
- D'ARMS, J. y JACOBSON, D. 2000, "Sentiment and Value", *Ethics*, 110: 722-48.
- D'ARMS, J. y JACOBSON, D. 2003, "The Significance of Recalcitrant Emotion (or, Anti-quasijudgmentalism)", *Royal Institute of Philosophy Supplement*, 52: 127-45.
- DAVIDSON, D. 1980, *Essays on Actions and Events: Philosophical Essays*, Clarendon Press.
- DE SOUSA, R. 1987, *The Rationality of Emotion*, MIT Press.
- DEIGH, J. 1994, "Cognitivism in the Theory of Emotions", *Ethics*, 104: 824-54.
- DEIGH, J. 2010, "Concepts of Emotions in Modern Philosophy and Psychology", P. GOLDIE (ed.), *The Oxford Handbook of Philosophy of Emotion*, Oxford University Press, 17-40.
- DENNETT, D. C. 1995, "Do Animals Have Beliefs?", H. ROITBLAT (ed.) *Comparative Approaches to Cognitive Sciences*, MIT Press.
- DEONNA, J. A. y TERONI, F. 2012, *The Emotions: A Philosophical Introduction*, Routledge.
- DEONNA, J. A. y TERONI, F. 2020, "Emotional Experience: Affective Consciousness And its role in Emotion Theory", U. KRIESEL (ed.), *Oxford Handbook of the Philosophy of Consciousness*, Oxford University Press.
- DRETSKE, F. 1988, *Explaining behavior: Reasons in a world of causes*, MIT Press.
- DUNN, R. 2006, *Values and the Reflective Point of View: On Expressivism, Self-knowledge and Agency*, Routledge.
- FINKELSTEIN, D. 2003, *Expression and the Inner*, Harvard University Press.
- FRIJDA, N. H. 2007, *The Laws of Emotion*, Lawrence Erlbaum Associates Publishers.
- GOLDIE, P. 2000, *The Emotions: A Philosophical Exploration*, Oxford University Press.
- GOLDIE, P. 2004, *On Personality*, Routledge.
- GOLDIE, P. 2010, "Love for a Reason", *Emotion Review*, 2 (1): 61-7.

- GORDON, R. 1987, *The Structure of Emotions: Investigations in Cognitive Philosophy*, Cambridge: Cambridge University Press.
- GRZANKOWSKI, A. 2017, "The Real Trouble with Recalcitrant Emotions", *Erkenntnis*, 82 (3): 641-51.
- GREEN, O. H. 1992, *The Emotions: A Philosophical Theory*, Dordrecht: Kluwer.
- GREENSPAN, P. 2011, "Craving the right: Emotions and moral reasons", C. BAGNOLI (ed.), *Morality and the Emotions*, Oxford University Press, 39-60.
- HELM, B. 2002, "Felt Evaluations: A Theory of Pleasure and Pain", *American Philosophical Quarterly*, 39: 13-30.
- HELM, B. 2009, *Love, Friendship, and the Self: Intimacy, Identification, and the Social Nature of Persons*, Oxford: Oxford University Press.
- HELM, B. 2015, "Emotions and Recalcitrance: Reevaluating the Perceptual Model", *dialectica*, 69: 417-33.
- JAMES, W. 1884, "What is an Emotion", *Mind*, 9: 188-205.
- KENNY, A. 1963, *Action, Emotion and Will*, Routledge.
- KIM, J. 1993, *Supervenience and Mind*, Cambridge University Press
- MARTÍNEZ, M. 2019, "A Mark of the Mental: In Defense of Informational Teleosemantics", *Notre Dame Philosophical Reviews*.
- McDOWELL, J. 1994, *Mind and World*, Harvard University Press.
- MITCHELL, J. 2021, *Emotion as Feeling Towards Value: A Theory of Emotional Experience*, Oxford University Press.
- NAAR, H. 2013, "A Dispositional Theory of Love", *Pacific Philosophical Quarterly*, 94: 3.
- NUSSBAUM, M. C. 2001, *Upheavals of Thought: The Intelligence of Emotions*, Nueva York: Cambridge University Press.
- PETTIT, P. 1991, "Realism and Response-Dependence", *Mind*, 100: 587-626.
- PINEDA, D. 2019, *Sobre las emociones*, Madrid: Cátedra.
- PRINZ, J. 2004, *Gut Reactions: A Perceptual Theory of Emotion*, Nueva York: Oxford University Press.
- RABINOWICZ, W. y RØNNOW-RASMUSSEN, T. 2004, "The Strike of the Demon: On Fitting Pro-Attitudes and Value", *Ethics*, 114: 391-423.
- ROBERTS, R. 2003, *Emotions: An Essay in Aid of Moral Psychology*, Cambridge: Cambridge University Press.
- SCARANTINO, A. 2010, "Insights and Blindspots of the Cognitivist Theory of Emotions", *British Journal for the Philosophy of Science*, 61(4): 729-68.
- SCARANTINO, A. y DE SOUSA, R. 2018, "Emotion", *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/entries/emotion/>>.
- SHEA, N. 2007, "Consumers Need Information: Supplementing Teleosemanantics with an Input Condition", *Philosophy and Phenomenological Research*, 75 (2): 404-35.

- SMITH, J. 2014, "Are Emotions Embodied Evaluative Attitudes?", *Disputatio*, 6(38): 93-106.
- SOLOMON, R. 2004, "Emotions, Thoughts, and Feelings: Emotions as Engagements with the World", R. SOLOMON (ed.), *Thinking about Feeling: Contemporary Philosophers on Emotions*, Oxford/Nueva York: Series in Affective Science, 76-88.
- SOLOMON, R. 1993, *The Passions: Emotions and the Meaning of Life*, Hackett Publishing Company.
- TAPPOLET, C. 2016, *Emotions, Values and Agency*, Oxford: Oxford University Press.
- VENDRELL FERRAN, I. 2018, "Emoción", *Enciclopedia de la Sociedad Española de Filosofía Analítica* (URL: <http://www.sefaweb.es/emocion/>)
- WIGGINS, D. 1987, *Needs, Values, Truth: Essays in the Philosophy of Value*, Oxford: Blackwell.
- WILLIAMS, B. 1976, "Moral Luck", *Proceedings of the Aristotelian Society*, 1: 115-35.
- WOLLHEIM, R. 1999, *On the Emotions*, New Haven y Londres: Yale University Press.