

QUADERNS

de Filosofia

VOL. VIII : I
VALÈNCIA, 2021

SOCIETAT DE FILOSOFIA DEL PAÍS VALENCIÀ

QUADERNS DE FILOSOFIA

Director: Sergi Rosell (Universitat de València)

Editors: José López González (Universitat Jaume I), Saúl Pérez González (Warsaw University of Technology)

Comitè editorial: Marc Artiga (Universitat de València), Ramón Feenstra (Universitat Jaume I), Antonio Gaitán Torres (Universidad Carlos III), Elsa González Esteban (Universitat Jaume I), Javier Gracia (Universitat de València), Elena Nájera (Universitat d'Alacant), Faustino Oncina (Universitat de València), Eduardo Rivera López (Universidad Torcuato di Tella), Pablo Rychter (Universitat de València), Jennifer Saul (University of Sheffield), José Zalabardo (University College London)

Comitè científic: María José Alcaraz (Universidad de Murcia), Jesús Alcolea (Universitat de València), Tyler Burge (University of California, Los Angeles), María Caamaño Alegre (Universidad de Valladolid), Enric Casaban (Universitat de València), Christine Chwaszcza (Universität zu Köln), Jesús Conill (Universitat de València), Adela Cortina (Universitat de València), Román de la Calle (Universitat de València), Antoni Defez (Universitat de Girona), Manuel García-Carpintero (Universitat de Barcelona), Domingo García-Marzá (Universitat Jaume I), Tomás Gil (Technische Universität Berlin), Antoni Gomila (Universitat de les Illes Balears), Tobies Grimaltos (Universitat de València), Susan Haack (University of Miami), Christopher Hookway (University of Sheffield), María Jiménez Buedo (UNED), Joan Llinares (Universitat de València), Genoveva Martí (Universitat de Barcelona – ICREA), Carlos Pereda (Universidad Nacional Autónoma de México), Diana Pérez (IFF-SADAF, CONICET, Universidad de Buenos Aires), Fernando Miguel Pérez Herranz (Universitat d'Alacant), Giovanna Pinna (Università del Molise), Xavier Serra (IES Joan Fuster, Sueca), Sergio Sevilla (Universitat de València), Ernest Sosa (Rutgers University), Agustín Vicente (Universidad del País Vasco – EHU), Vanessa Vidal Mayor (Universitat de València), Nicholas White (University of California, Irvine)

Edita: Societat de Filosofia del País Valencià

eISSN: 2341-3042

Disseny i seguiment editorial: Antoni Domènech

© de l'edició, Societat de Filosofia del País Valencià

© dels articles, els seus autors

<http://ojs.uv.es/index.php/qfilosofia>

Redacció: Societat de Filosofia del País Valencià

Universitat de València — Facultat de Filosofia i Ciències de l'Educació

Departament de Filosofia

Av. Blasco Ibáñez, 30-5a. 46100 València

e-mail: sfpv@uv.es

<http://www.uv.es/sfpv/>

QUADERNS DE FILOSOFIA és una revista acadèmica dirigida als professionals de la filosofia de periodicitat semestral. Publica articles de temàtica filosòfica de qualitat científica i té un interès especial en la didàctica de la filosofia.

Cada número ordinari té almenys tres parts: 1. Articles d'investigació: contribucions originals pertanyents als diversos àmbits de la recerca filosòfica; 2. L'anomenada "Brúixola filosòfica", feta d'articles que presenten un determinat tema o debat filosòfic; i 3. Documents, entrevistes, notes de discussió i ressenyes.

Tots els articles originals són sotmesos a una revisió cega anònima per experts en l'àrea corresponent.

La revista es publica en versió electrònica d'accés obert (Open Journal System).

La revista és inclosa en alguns dels índexs i bases de dades de major prestigi internacional en filosofia i humanitats, són CIRC, Dialnet, DOAJ, Dulcinea, ERIH PLUS, ÍNDICES-CSIC, Latindex, MIAR, Philosopher's Index, REBIUN, REDIB, i l'equip editorial treballa per a la inclusió en d'altres, en especial FECYT i SCOPUS.

Normes de publicació

Els articles poden estar escrits en català, castellà i anglès, i no han d'excedir de les 10.000 paraules. Han de dur títol i incloure un resum (de fins a 150 paraules) i unes paraules clau en la llengua de l'article i en anglès. Els manuscrits s'han de presentar de forma anònima per tal de garantir-ne el dictamen cec; és responsabilitat de l'autor eliminar qualsevol autoreferència.

També es poden proposar ressenyes de llibres publicats en els darrers tres anys, amb una extensió màxima de 2.500 paraules, llevat de casos excepcionals prèviament acordats amb la direcció de la revista. Les ressenyes no tindran títol ni notes. Com a encapçalament cal consignar les dades completes del llibre ressenyat, inclòs l'ISBN. Cal recordar que una ressenya no és un resum, sinó un examen crític del llibre ressenyat.

Les normes d'estil poden consultar-se en la versió electrònica de la revista, però en tot cas sols s'exigirà el seu compliment en la versió final de l'article o ressenya, si se n'accepta la proposta.

ÍNDIX

Articles

- “Las representaciones conceptuales en la ciencia cognitiva”
Mariela DESTÉFANO (Universidad de Buenos Aires - CONICET) ... 11
- “(Not) Hosting Refugees in the EU: Philosophical Perspectives
on the 2015-2016 Crisis”
BOSCO CORRALES TRILLO
(Universidad Católica de Valencia “San Vicente Mártir”) 37

Brúixola filosòfica

- “El debat sobre l’estructura causal de la teoria evolutiva:
lleis, forces i analogies”
VÍCTOR J. LUQUE (Universitat de València) 63

Documents

- “Presentación”
Alberto OYA (Universidade Nova de Lisboa) 101
- “Definiciones persuasivas”
Charles Leslie STEVENSON 105

Notes de discussió i ressenyes

- Faustino ONCINA (ed.), *Crítica de la Modernidad. Modernidad de la crítica*, 2019.
Per Joan GALLEGRO MONZÓ (Universitat de València) 129

Articles

MARIELA DESTÉFANO

Universidad de Buenos Aires

Consejo Nacional de Investigaciones Científicas y Tecnológicas

Las representaciones conceptuales en la ciencia cognitiva¹

Conceptual Representations in Cognitive Science

Recibido: 13/9/2019. Aceptado: 22/4/2021

Resumen: En el marco de la ciencia cognitiva se ha polemizado acerca de la manera en la que el enfoque enraizado de la cognición pueda compatibilizarse con el enfoque clásico con el fin de explicar nuestras capacidades conceptuales. Sin embargo este debate metodológico no ha sido acompañado de una elucidación de la noción de “representación conceptual” tal como venía siendo entendida al menos en la filosofía y psicología cognitivas. El objetivo de este trabajo es analizar en que consiste una representación conceptual con el fin de aportar claridad al debate sobre las capacidades conceptuales enraizadas. Desarrollaré y evaluaré la idea de que una representación es *conceptual* si puede combinarse sistemáticamente para formar nuevas estructuras representacionales y si se puede utilizar en tareas psicológicas con independencia del estímulo.

Abstract: In cognitive science, it is an open debate whether grounded cognition might be compatible with traditional views of cognition. However, as far as I am aware, this methodological debate has not been accompanied by an elucidation of the notion of “conceptual representation” as it has been understood in philosophy and cognitive psychology. The aim of this paper is to offer an elucidation of this sort. I will develop and evaluate the idea that a representation is *conceptual* when it can be systematically combined to form new representational structures and when it can be used in psychological tasks regardless of the stimulus.

¹ Quisiera agradecer a los evaluadores de la revista por los fructíferos comentarios que me han permitido mejorar la presentación de este artículo.

Palabras clave: capacidades conceptuales, sistematicidad, independencia del estímulo, cognición enraizada, psicología cognitiva.

Keywords: conceptual abilities, systematicity, independence of stimulus, grounded cognition, cognitive psychology.

I. INTRODUCCIÓN

EN EL MARCO DE LA CIENCIA COGNITIVA se ha propuesto que nuestras capacidades conceptuales estarían enraizadas en otras capacidades tales como las sensitivas, perceptivas, motoras y afectivas (BARSALOU 2016; PECHER 2013; MARTIN 2016; GOLDINGER 2016; MAHON 2015). Según Weber y Vosgerau (2018), “enraizado” puede interpretarse tanto en términos de condiciones de adquisición como en términos de condiciones de constitución. En el primer caso, una habilidad cognitiva A está enraizada en la base de B si B es necesario para adquirir A. En el segundo caso, una habilidad cognitiva A está enraizada en B si B es necesario para poseer la habilidad A. Muchos desarrollos en el enfoque enraizado se focalizan en el rol de las simulaciones en la cognición, entendiendo por simulación la reactivación de los estados perceptivos, motores e introspectivos adquiridos durante la experiencia (BARSALOU 2008). En el transcurso de las experiencias (e.g. usar una silla) el cerebro captura información a través de las distintas modalidades perceptivas y la integra en una representación multimodal almacenada en la memoria. Luego, cuando en aquellas tareas que demanden representar un concepto (e.g. SILLA), estas representaciones multimodales capturadas durante la experiencia son reactivadas.

Más allá de las particularidades que tengan las distintas propuestas enraizadas, todas se contraponen al enfoque clásico según el cual a las capacidades conceptuales subyacen representaciones que son independientes de los sistemas modales del cerebro para la percepción, la acción, afectividad e introspección (BARSALOU 2010). Sin embargo, es improbable que el enfoque enraizado tenga éxito sin el compromiso con elementos propios del enfoque clásico según el cual las representaciones conceptuales son simbólicas (BARSALOU 2008). De este modo, uno de los grandes objetivos explicativos de los defensores de la cognición enraizada es encontrar el modo de compatibilizar ambos enfoques. En este sentido, en los últimos años ha tenido lugar un debate metodológico (que denominaré “debate enraizado”) el cual es prolífico y ha generado una pluralidad de posiciones respecto de cómo entender la integración entre en enfoque enraizado y el clásico.

Por un lado, una de las maneras de entender dicha integración es considerar que las representaciones simbólicas y las simulaciones operan en una

misma arquitectura cognitiva. Este es el caso de propuestas teóricas como las de Dove (2009), Louwerse (2012) o Zwan (2014) para quienes los procesos conceptuales involucran ambos tipos de representaciones. Por ejemplo, Zwan ha propuesto que las representaciones simbólicas se utilizan en tareas que demandan mayor nivel de abstracción mientras que las simulaciones se utilizarían en las interacciones concretas con el ambiente. De acuerdo con esta manera de entender la integración ambos tipos de representaciones pertenecen al mismo nivel funcional. Frente a ello, autores como Barsalou (2010) entienden que las simulaciones podrían constituir una propuesta de implementación neural de una arquitectura simbólica clásica. Aunque su teoría pertenece originariamente a la psicología cognitiva, Barsalou entiende que el enfoque enraizado se acomoda “naturalmente” a la investigación cerebral “debido a que toma en consideración las modalidades perceptivas” (2008, 635). En este caso, ambos enfoques se integrarían pero desde diferentes niveles explicativos, debido a que el enfoque enraizado constituiría una implementación cerebral de las propiedades funcionales que tendría una arquitectura clásica de representaciones simbólicas.

Hasta lo que sé, este debate metodológico no ha sido acompañado de una elucidación de la noción de “representación conceptual” tal como venía siendo entendida al menos en la filosofía y psicología cognitivas. Por esta razón, el objetivo de este trabajo es analizar en que consiste una representación conceptual con el fin de aportar claridad conceptual al debate sobre las capacidades conceptuales enraizadas. Hacer jugar en este debate un criterio de mínima de representación conceptual sustentada por la literatura sobre el tema contribuiría, en parte, a clarificar algunos elementos de juicio a favor de una u otra posición en el debate sobre cognición enraizada. Pues, habiendo elucidado cuáles son las condiciones para una representación conceptual, el hecho de que tanto representaciones simbólicas como simulaciones cumplan con dichas condiciones podría significar que ambos tipos de representaciones operarían conjuntamente en una misma arquitectura cognitiva aunque haciendo diferentes aportes en tareas cognitivas diferentes. Por otro lado, un análisis conceptual de las simulaciones que muestre que no cumplen con ciertas condiciones mínimas de conceptualidad podría indicar o bien que pertenecen a un nivel de implementación neuronal o bien que hacen un aporte a las arquitectura cognitivas simbólicas pero en carácter de representaciones no conceptuales. De hecho, más adelante abriré la posibilidad de considerar a las simulaciones o bien como representaciones no conceptuales o bien como entidades neuronales. Es cierto que profundizar en estos enfoques es algo que excede este trabajo y que demanda evidencia empírica. Lo importante en este contexto es que ambas interpretaciones tendrían lugar de momento que las simulaciones sólo cumplirían parcialmente con los criterios de conceptualidad presentados.

Existe un consenso filosófico en sostener que los contenidos de los estados de actitud proposicional (por ejemplo, lo creído en el caso de las creencias o lo deseado en el caso de los deseos) están constituidos por conceptos, en la medida en que estar en un estado de actitud proposicional requiere la posesión de los mismos. Por ejemplo, un individuo no puede creer (ni formular el juicio de) que el gato es perezoso sin poseer, al menos, los conceptos GATO y PEREZOSO. Es cierto que la literatura sobre conceptos involucra una diversidad de discusiones planteadas tanto en el plano filosófico como psicológico. En la filosofía ha sido usual considerar a los conceptos como entidades abstractas que carecen de especificaciones psicológicas (FREGE 1892). Frente a ello están los debates psicológicos sobre conceptos que en su mayoría han asumido una ontología de particulares mentales. Como el objetivo de este trabajo es focalizarse en los desarrollos de las ciencias cognitivas donde suele adherirse a una perspectiva de particulares mentales en este trabajo se adoptará la idea de que los conceptos son representaciones mentales, dejando de lado planteos que desnudan a los conceptos de todo factor psicológico. En este marco el contenido de los estados de deseos y creencias está constituido por conceptos, entendidos como representaciones que son particulares mentales instanciados físicamente, que poseen propiedades tales como activarse y procesarse de manera independiente del estímulo y combinarse satisfaciendo el requisito de generalidad (EVANS 1982).² En este trabajo desarrollaré y evaluaré la idea de que una representación es conceptual si puede combinarse sistemáticamente según el requisito de generalidad para formar nuevas estructuras representacionales y si se puede utilizar en tareas psicológicas (e.g. categorización, razonamiento práctico, etc.) con independencia del estímulo. En la sección 2 haré explícito que rescato una caracterización funcional de las representaciones conceptuales. En la sección 3 presentaré la idea de que los conceptos son representaciones independientes del estímulo y diré algunas palabras relativas a la posibilidad de que la noción de “simulación” respete este criterio de conceptualidad. En la sección 4 desarrollaré la idea de que son representaciones que se combinarían sistemáticamente y también evaluaré en qué sentido las simulaciones no se combinarían de esta manera. En ambas secciones focalizo en la noción de “simulación perceptiva” propia de Barsalou (1999A; 1999B) debido a la centralidad que tiene dicha noción en el enfoque enraizado. Si bien es un concepto

² Cabe aclarar que existe una significativa interacción entre los enfoques filosóficos y psicológicos acerca de los conceptos (LAURENCE y MARGLIS 1999). Lingüistas y psicólogos se han inspirado en los aportes wittgenstorianos sobre los parecidos de familia o en la distinción fregeana entre sentido y referencia, o en las tesis extenalista y esencialistas de Kripke y Putnam para el planteo de sus teorías empíricas. También muchos debates filosóficos se han enriquecido con los aportes de la psicología acerca de las tareas de categorización. Esta retroalimentación ha permitido incorporar aportes de filósofos como Evans en el marco de las ciencias cognitivas.

que pertenece a la psicología cognitiva es compatible con otras propuestas enraizadas de la ciencia cognitiva como por ejemplo la de Prinz (2002). En las conclusiones retomaré los lineamientos generales presentados en las anteriores secciones e indagaré brevemente sobre las razones por las cuales el debate enraizado se vería enriquecido con este análisis de las condiciones de mínima para las representaciones conceptuales.

2. EL ROL FUNCIONAL DE LOS CONCEPTOS

Generalmente, el criterio para determinar qué tipo de entidades son los conceptos apela al rol funcional que este tipo de representaciones debería satisfacer en nuestra vida mental. Hacer una caracterización funcional de una entidad significa caracterizarla en términos de sus propiedades funcionales. Las propiedades funcionales son las que describen lo que hace una entidad, cuál es su rol en un sistema más amplio (WEISKOPF ms.). Así, caracterizar funcionalmente a los conceptos significa dar una descripción de cuáles son los roles que debe satisfacer una entidad para ser concepto: “algo es un concepto en virtud de la función que desempeña en un sistema cognitivo” (LALUMERA 2010, 217). Una vez que se especifican estos roles, todas aquellas entidades que los satisfacen serían conceptos. Los roles funcionales que deberían satisfacer los conceptos varían según las particularidades de cada propuesta teórica. Esta especificación funcional de las representaciones conceptuales pretende rescatar lo que, en un sentido general, filósofos y psicólogos han entendido por “concepto”.

Por ejemplo, Prinz considera que los conceptos tienen que cumplir el rol de constituirse como representaciones que están bajo control endógeno: “Sugiero que los conceptos son espontáneos: son representaciones que pueden estar bajo control endógeno del organismo” (2002, 197). Una representación bajo control endógeno es activada y manipulada por el sujeto que la posee. En este sentido, una representación bajo control endógeno es consciente para el sujeto que la posee, teniendo en cuenta que para Prinz (2004) la posibilidad de control del organismo involucra un acceso consciente a la representación. Asimismo, una representación bajo control endógeno se produce voluntariamente, en tanto que “voluntariamente se pueden formar pensamientos usando conceptos” (PRINZ 2004, 46). Pero decir que los conceptos son representaciones manipuladas de manera consciente y voluntaria, dificultaría la posibilidad de afirmar que hay sistemas de procesamiento temprano de la información (como el lingüístico) que operan sobre conceptos. Estos sistemas de procesamiento operarían sobre representaciones subpersonales las cuales no pueden

ser manipuladas consciente y voluntariamente por un sujeto. Por lo tanto, la razón para dejar de lado la caracterización de Prinz es que no constituye una condición de mínima que pueda aplicarse a la amplitud de casos de conceptualidad (que va del nivel personal al subpersonal) que existe en la psicología y la filosofía de la ciencia cognitiva.

Weiskopf (ms.), por su parte, asigna a los conceptos un rol funcional diferente:

La función de los conceptos es permitir a una criatura que los tiene trascender las maneras de representar el mundo que sus otros sistemas cognitivos le proporcionan, especialmente, los sistemas sensorio-motores (13)

Entendidos de esta manera, los conceptos son representaciones abstractas que representan propiedades que no podrían ser representadas por los sistemas sensorio-motores:

Hay muchas cosas que no pueden ser representadas por nuestros sistemas dedicados para la percepción, [...] y el lenguaje, etc. Los espacios de Hilbert, los objetos matemáticos, no tienen manifestaciones perceptivas y, seguramente, exceden los poderes de nuestra facultad numérica. La propiedad de ser una ilusión del Infierno de Dante es algo que la facultad del lenguaje no puede representar (WEISKOPF ms., 12)

Así, los conceptos también serían representaciones abstractas en el sentido de que pueden representar las mismas propiedades representadas por los sistemas sensorio-motores pero de manera diferente. Por ejemplo, puede haber conceptos de cualidades perceptivas que no las representan “bajo un modo de presentación perceptivo” (WEISKOPF ms., 13). Considero que el criterio de abstracción es correcto, sin embargo, según mi entender, no parece ser un criterio de mínima que hace que una representación sea conceptual. Más adelante adoptaré como una de las condiciones de mínima para la conceptualidad la independencia del estímulo. Podría especularse que el carácter abstracto de las representaciones conceptuales estaría posibilitado por su independencia del estímulo. Aunque no me detendré sobre esta cuestión, puede pensarse que una representación es abstracta en tanto que, en algún sentido, puede instanciarse sin que sea activada por un estímulo perceptivo. De este modo, la abstracción no sería un criterio de mínima para la conceptualidad sino que dependería, en parte, del de la independencia del estímulo. Además, considerar a la abstracción como una condición de mínima para la conceptualidad significaría alejarse de la literatura psicológica y filosófica de raigambre enraizada.

Otra manera de entender el rol funcional de los conceptos es a partir de la siguiente caracterización que hace Machery:

Un concepto de x es un cuerpo de conocimientos sobre x que está almacenado en la memoria de largo plazo y se usa por *default* en los procesos subyacentes a la mayoría de las competencias cognitivas superiores, cuando estos procesos resultan en juicios sobre x (2009, 12)

Según esta caracterización, los conceptos son representaciones almacenadas en la memoria de largo plazo. Sin embargo, muchos filósofos y psicólogos defienden la posibilidad de que los procesos subyacentes a las competencias cognitivas superiores, de las que habla Machery, operen sobre representaciones temporalmente construidas en la memoria de trabajo (BARSALOU 1999A; PRINZ 2002; WEISKOPF ms.). Tal como sucede con la propuesta de Weiskopf, asumir este criterio como condición mínima de conceptualidad significaría alejarse de muchas de las propuestas enraizadas que serán relevantes en este trabajo.

Desarrollaré dos criterios funcionales de mínima para la conceptualidad que se distancian de los propuestos por los autores anteriores. Según estos criterios, ciertas representaciones son conceptuales porque se emplean en una variedad de tareas cognitivas donde suele atribuirse conceptos con independencia del estímulo y se pueden combinar sistemáticamente para formar nuevas estructuras. La primera condición es la que han propuesto sobre todo los psicólogos y la segunda condición es la que han propuesto los filósofos. El motivo que llevaría a establecer condiciones de conceptualidad que han impuesto tanto psicólogos como filósofos es que esta noción de “concepto” sea útil desde un punto de vista filosófico y empírico. Es decir que estas condiciones no son arbitrarias sino que, por el contrario, suelen aparecer en las propuestas que caracterizan al contenido conceptual. De esta manera, esta noción de “concepto” es descriptiva porque rescata lo que la mayoría de los psicólogos y filósofos entienden por “concepto”, aunque no haya mucho acuerdo entre ellos ni al interior de cada disciplina.

En psicología cognitiva, no se utiliza un vocabulario unificado para referirse a conceptos. Algunos usan términos que provienen de la inteligencia artificial, como “frames” o “scripts”, otros usan “conceptos” y “categorías” de manera intercambiable (SMITH y MEDIN 1981). Y las diferencias en el tratamiento de los conceptos van más allá de lo meramente terminológico. Existen muchas teorías psicológicas que se diferencian en el tipo de información que consideran que constituye la estructura del contenido de un concepto. Según algunas de estas teorías, los conceptos pueden ser prototipos (ROSH 1978),

ejemplares (ESTES 1994) o teorías (REHDER 2003). Pero a pesar de estas diferencias, todas las investigaciones psicológicas sobre representaciones conceptuales acuerdan en sostener que la información conceptual es la que interviene en tareas como categorización, comprensión y producción lingüística (PICCININI y SCOTT 2006).

La condición filosófica acerca de la conceptualidad, según la cual los conceptos pueden recombinarse de manera sistemática para formar otras estructuras, es lo que rescata Evans (1982) con el requisito de generalidad. El requisito de generalidad surge al prestar atención a la estructura de los pensamientos. Los pensamientos tienen una estructura que refleja las relaciones entre los constituyentes de las oraciones del lenguaje.

3. CONCEPTUALIDAD E INDEPENDENCIA DEL ESTÍMULO

Para caracterizar a los conceptos “tenemos que atender al rol que tienen en las teorías psicológicas contemporáneas” (PICCININI y SCOTT 2006, 295). No es controversial sostener que el contenido conceptual es de vital importancia en tareas cognitivas que Piccinini y Scott (2006) identifican en la siguiente lista:

- Discriminación: responder de manera diferenciada a objetos, propiedades y eventos que tienen algo saliente en común.
- Categorización: asociar de manera arbitraria un *input* con una respuesta, incluyendo la habilidad de producir una etiqueta apropiada (e.g., una palabra) en respuesta a una clase de *inputs*, así como también elegir el objeto o comportamiento apropiado en respuesta a una etiqueta dada como estímulo (ROSH 1999).
- Inferencias no lingüísticas: establecer inferencias sobre clases de objetos en el mundo. Puede incluir el razonamiento deductivo (BRAINE *et al.*, 1984), analógico, (GENTNER 1983), razonamientos prácticos (GREENE 2009), entre otros.
- Inferencias lingüísticas: establecer inferencias que incluyen palabras (e.g., la inferencia que va de “rojo” a “coloreado”) e inferencias entre oraciones en virtud de su gramática.
- Comprensión y producción lingüística: procesar las palabras y las oraciones ante un *input* lingüístico de habla o escritura (ELLIS y YOUNG 1992).³

³ Los componentes no conceptuales también intervienen en algunas de las tareas que aparecen en esta lista. Ver nota 7 en Piccinini y Schott (2006).

Esta caracterización sigue de cerca la noción de “concepto” que puede extraerse de la psicología cognitiva. Aunque es difícil identificar una noción unívoca de concepto compartida por las diferentes teorías psicológicas (MACHERY 2009), en un sentido general, todas las teorías de conceptos estarían dispuestas a sostener que estos son cuerpos de información usados en tareas como categorización, deducción, inducción, etc. Los psicólogos suelen entender que, por ejemplo, las tareas en las que el sujeto tiene que realizar inferencias no lingüísticas y lingüísticas requieren la participación de conceptos. Tomemos la siguiente inferencia: “Si se da X e Y , entonces no se da Z . Por lo tanto, si se da Z , o no se da X o no se da Y ” (MONTI *et al.* 2009). Es una inferencia deductiva en la que se llega a la conclusión haciendo uso de reglas lógicas y utilizando conceptos como X , Y y Z . Tomemos ahora la siguiente inferencia lingüística: “Era X lo que Y vio tomar a Z . Por lo tanto, Z fue visto por Y tomando X ” (MONTI *et al.* 2009). En este caso, la inferencia es lingüística porque se realiza en virtud de la gramática de las oraciones y no en virtud de reglas lógicas. También en esta oportunidad la inferencia se establece utilizando conceptos como X , Y y Z .

Las representaciones que subyacen a las tareas cognitivas como las anteriores son conceptuales en el sentido de que satisfacen lo que Millikan (2000) denominó “requisito de distancia”. Desde esta perspectiva, una representación es conceptual en tanto que puede activarse y procesarse con independencia del estímulo. Esto quiere decir que aquello en el mundo que activaría el concepto no es necesario para su ocurrencia, lo cual plantea una separación entre la representación y el *input* de tal manera que puede darse la ocurrencia de la primera sin la ocurrencia del segundo (CAMP 2009). Este criterio para la conceptualidad concierne a la relación que tiene una representación con aquello que la instancia: una representación sería conceptual cuando puede activarse y procesarse con independencia de los estímulos del mundo. Por supuesto que hay circunstancias en las que, por ejemplo, PERRO se activa y procesa en caso de que haya algún perro como estímulo distal. Pero si, en ciertas tareas cognitivas, PERRO puede activarse y procesarse sin la necesidad de que haya un perro como estímulo, entonces podría pensarse que esta representación es conceptual.

Una representación es independiente del estímulo o bien porque el estímulo no está presente, o bien porque el estímulo está presente pero podría no intervenir en la activación y el procesamiento de la representación. En este sentido, la independencia del estímulo se manifiesta de dos maneras, tanto en circunstancias en las que efectivamente no está el estímulo pertinente como en circunstancias en las que puede estar el estímulo pertinente. Quisiera explorar con algún detalle ambas formas de la independencia del estímulo. En primer lugar, el estímulo de la representación no está presente en el sentido de no encontrarse en el ambiente circundante al organismo que posee el concepto. Las

representaciones cuyos estímulos se encuentran en ambientes lejanos se usan en cierta clase de razonamiento espacial. Imaginemos el siguiente razonamiento espacial. El objeto *A* está a la distancia *X* de donde estoy. El objeto *B* está cerca del objeto *A*, por lo tanto el objeto *B* está a la distancia *X* de donde estoy. Este razonamiento exige representaciones del objeto *A* y el objeto *B*, cuyos estímulos no se encuentran en el ambiente inmediato dado que están a la distancia *X*.

A su vez, el estímulo de una representación puede no estar presente en el sentido de que dicho estímulo no existe. Los razonamientos prácticos de toma de decisión exigen la utilización de representaciones independientes del estímulo en este sentido. En un razonamiento de toma de decisión, un individuo elige uno entre dos o más posibles cursos de acción. Decidir no es simplemente seleccionar uno de estos cursos de acción: “una genuina toma de decisión involucra la selección entre diferentes cursos de acción basada en las consecuencias probables que esos cursos de acción tendrían” (BERMÚDEZ 2003, 124). Es decir que una toma de decisión involucra algún tipo de proceso de comparación entre diferentes cursos de acción y toma en cuenta sus posibles consecuencias. Tal como sostiene Bermúdez: “esto requiere representar los diferentes resultados potenciales, representar la utilidad que tiene cada uno de ellos, y estimar la probabilidad de que cada resultado ocurra” (2003, 125). Lo importante a destacar es que la representación de los diferentes resultados potenciales o la representación de la utilidad que tienen las consecuencias de los cursos de acción que se evalúan en la toma de decisión se activan aun cuando en el mundo no haya un estímulo distal que produzca su instanciación.

En segundo lugar, una representación también es independiente cuando el estímulo puede estar presente en el ambiente circundante al organismo que posee el concepto y aun así, no activa la representación en cuestión. Este tipo de representaciones independientes del estímulo son las necesarias para tener lo que denomino “pensamientos solipcistas”. Es concebible la posibilidad de que un sujeto piense que los collies son bellos perros cuando en su ambiente circundante hay un collie, y sostener al mismo tiempo que este collie no ha sido necesario para la activación y el procesamiento de COLLIE por parte del sujeto. Este concepto ha sido autogenerado en el sentido de que el estímulo, aun cuando estaba presente, no lo ha provocado.

A diferencia de las representaciones independientes del estímulo, las representaciones dependientes del estímulo necesitan la presencia del estímulo pertinente para su activación y procesamiento. Son representaciones en las que hay una separación débil entre la representación y el estímulo dado que no puede darse la primera sin que el segundo sea el caso. Un ejemplo en el cual sería necesario postular representaciones dependientes del estímulo es cuando, en el reino animal, los animales machos evalúan a otros macho en una situa-

ción de competencia (BERMÚDEZ 2003). Mediante la sola inspección del macho contrincante “el animal podría ver que pelear es una respuesta apropiada o inapropiada” (BERMÚDEZ 2003, 121). Sus posibilidades de acción resultan de la sola inspección de las condiciones percibidas en el ambiente. El ambiente provee toda la información necesaria (e.g., un macho muy pequeño) para explicar la respuesta conductual de un animal (e.g., iniciar una pelea). Si se hace la concesión de que este tipo de conductas debe explicarse atribuyendo representaciones a los animales, estas representaciones deberían entenderse como dependientes del estímulo.

Otro ejemplo en el cual sería necesario atribuir representaciones dependientes del estímulo sería en el caso de los pacientes con síndrome de dependencia del estímulo. Este síndrome es producto de lesiones en el lóbulo frontal y se caracteriza por el hecho de que los pacientes se desempeñan en sus tareas cognitivas atendiendo de manera exclusiva a los estímulos ambientales inmediatos (LHERMITTE 1986). Un ejemplo de ello es el paciente A.D. quien asumía diferentes roles según el contexto en el que estaba (CONCHIGLIA *et al.* 2007). En sus discusiones con los psicólogos, A.D. sostenía ser un psicólogo y actuaba como tal. En el hospital, frente a los médicos, A.D. decía ser médico así como también, en ocasión de una visita a un edificio en construcción, A.D. sostenía ser ingeniero y asumía dicho rol. Esta singular conducta podría explicarse atribuyendo a A.D. representaciones que dependen del estímulo. Este paciente se representa ser un psicólogo, médico o ingeniero atendiendo de manera exclusiva a la información que le brinda el ambiente en el que está.

Es importante destacar que la distinción entre dependencia e independencia del estímulo es independiente respecto de la determinación internalista o externalista del contenido. Son concebibles todas estas opciones:

- (i) Una representación dependiente del estímulo cuyo contenido está determinado de manera externalista.
- (ii) Una representación dependiente del estímulo cuyo contenido está determinado de manera internalista.
- (iii) Una representación independiente del estímulo cuyo contenido está determinado de manera externalista.
- (iv) Una representación independiente del estímulo cuyo contenido está determinado de manera internalista.

La pregunta previa que surge es ¿por qué podemos admitir las opciones que van de (i) a (iv)? Porque una cosa es tomar partido respecto de cómo se activa y procesa una representación, y esto se relaciona con la presencia o ausencia del estímulo pertinente para la activación y procesamiento de la

representación, y otra cosa es tomar posición respecto de la manera en que se individúa el contenido de la representación, y esto se relaciona estrictamente con la semántica de la representación. Quiero exponer esta idea tomando en consideración, sobre todo, las opciones (ii) y (iii).

La opción (ii) asume la posibilidad de una representación dependiente del estímulo cuyo contenido está determinado de manera internalista. Por ejemplo, una representación táctil que se ha activado a causa de la presencia de un objeto rugoso es dependiente del estímulo porque hubo un objeto en el mundo que dio lugar al procesamiento de dicha representación. Pero establecer esta relación entre el estímulo de la representación y la representación no obliga desde un punto de vista conceptual a adoptar una semántica externalista. El contenido de esta representación no necesariamente está determinado por el objeto rugoso en el mundo que causó la representación, también puede concebirse que este contenido esté determinado por el rol que tiene en relación a otras representaciones del sistema cognitivo, por ejemplo.

La opción (iii) es la que suele adoptar la mayoría de los filósofos y asume la posibilidad de una representación independiente del estímulo cuyo contenido está determinado de manera externalista. Por ejemplo, la representación conceptual PERRO sería independiente del estímulo en la medida en que puede darse la activación del concepto sin que haya ningún estímulo pertinente en el mundo que lo active. Sin embargo, aunque la representación no depende del estímulo para activarse, aun así se puede sostener que el contenido de PERRO está determinado por la propiedad de la *perritud* en el mundo.

Luego de haber presentado la primera condición de mínima para una representación conceptual la pregunta que surge es si las “simulaciones perceptivas”, las cuales propone Barsalou (1999A; 1999B) en el enfoque enraizado, son independientes del estímulo. Barsalou (1999B) define a las simulaciones como activaciones top-down de áreas sensorio-motoras que reactiva experiencias perceptivas. Éstas son creadas por mecanismos específicos denominados “simuladores” los cuales permiten al sistema cognitivo construir representaciones de “una entidad o eventos en su ausencia” (BARSALOU 1999, 586). Por ejemplo, en la construcción de la simulación de un auto, la atención selectiva se focaliza en diferentes aspectos del objeto los cuales quedan organizados en un registro perceptivo espacialmente organizado. La información perceptiva integrada en este registro servirá posteriormente para simular nuevas experiencias coherentes del objeto lo cual permitirá al sistema cognitivo anticiparse a nuevos encuentros con autos. El objeto simulado tiene aspectos multimodales de la experiencia que quedan incorporados en la múltiple información perceptiva (e.g. visual, táctil, auditiva) de las simulaciones. Dada esta caracterización de las simulaciones puede pensarse que en términos de cómo se construyen las mismas no

serían independientes del estímulo. Pues, el ambiente contribuye causalmente en la genética de las simulaciones. Los diferentes aspectos de la experiencia reclutados por la atención selectiva tienen un rol central en la constitución de las mismas. En este sentido, las simulaciones carecerían de cualquier versión de independencia del estímulo de momento que son representaciones enraizadas en tanto que su adquisición involucra información perceptiva extraída del ambiente (WEBER y VOSGERAU 2018).

Sin embargo existe un sentido interesante en el cual las simulaciones serían independientes del estímulo. Barsalou (2008) explicita que las simulaciones de hecho proceden independientemente de las experiencias sensorio-motoras concretas. La idea central de enfoque enraizado es que los mismos procesos perceptivos, afectivos y motores involucrados en la interacción con objetos y eventos, se activan cuando razonamos, comprendemos y pensamos en general acerca de esos objetos y eventos. Las tareas conceptuales (e.g. tareas de verificación de propiedades, razonamiento espacial, razonamiento abstracto, comprensión de palabras, entre otras) involucrarían simulaciones almacenadas en la memoria de trabajo (BARSALOU 2008). La particularidad que tienen las tareas conceptuales es que una vez superado el análisis perceptivo inicial los mecanismos que usan las simulaciones operan de manera independiente a cualquier estímulo. Por ejemplo, en una tarea de verificación de propiedades, donde los sujetos tienen que decidir si una propiedad (e.g. CUERNOS) pertenece a una categoría (CABALLOS), la utilización de simulaciones de los conceptos y sus propiedades que permiten desarrollar la tarea son procesadas con independencia de la ocurrencia real de objetos como cuernos o caballos. De modo que las simulaciones serían representaciones independientes del estímulo que se activan por el solo funcionamiento de los procesos conceptuales.

4. CONCEPTUALIDAD Y REQUISITO DE GENERALIDAD

La segunda caracterización de las representaciones conceptuales es un requisito filosófico que surge de la literatura analítica. El contenido conceptual es aquel cuyos constituyentes se pueden combinar sistemáticamente para formar nuevas estructuras representacionales. Esta idea fue presentada por Evans (1982, 104) en lo que denomina el “requisito de generalidad”:

Si a un sujeto se le atribuye el pensamiento de que a es F , entonces debe tener los recursos conceptuales para tener el pensamiento de que a es G , para cualquier propiedad de ser G de la cual tiene una concepción. Esta es la condición que llamo “El requisito de generalidad”

Los filósofos suelen sostener que cierto tipo de pensamientos están estructurados. Los pensamientos *Helena es salvaje* y *Aristides es mimoso*. El contenido de estos pensamientos está constituido por conceptos como HELENA y SALVAJE, en el primer caso, y ARÍSTIDES y MIMOSO, en el segundo caso. Estos pensamientos están estructurados en tanto que su contenido, constituido por conceptos, responde a las formas Fa y Gb , en donde a es HELENA, F es SALVAJE, b es ARÍSTIDES, y G es MIMOSO. Lo que plantea Evans (1982) con el requisito de generalidad es que si un sujeto tiene los pensamientos estructurados Fa y Gb , entonces no habría ningún impedimento conceptual para que tenga los pensamientos Ga y Fb . La idea es que la posibilidad de tener nuevos pensamientos está dada por la recombinación de los constituyentes de otros pensamientos del sujeto. Así, el pensamiento Fa sería la intersección entre dos series, la de Fa, Fb, \dots , y la de Fa, Ga, \dots . Asimismo, el pensamiento Gb , sería la intersección entre dos series, la de Ga, Gb, \dots , y la de Fb, Gb, \dots (EVANS 1982, 104).

Evans extrae la idea de que tener ciertos pensamientos nos compromete con la posibilidad de tener otros pensamientos a partir de lo que sucede en el caso de las oraciones del lenguaje natural. En este caso, si un sujeto comprende “ Fa ” y “ Gb ”, entonces debemos comprometernos con la idea de que el sujeto también comprenderá “ Fb ” y “ Ga ”. Evans establece una analogía según la cual, si las oraciones del lenguaje natural y los pensamientos comparten la característica de estar estructurados, y si la comprensión de ciertas oraciones involucra la comprensión de otras oraciones, entonces tener ciertos pensamientos involucra tener otros pensamientos. Los límites que Evans ve en esta analogía es que mientras que los pensamientos están esencialmente estructurados, las oraciones del lenguaje natural no estarían esencialmente estructuradas.⁴

Pero Evans (1982, 101) va más allá y sostiene que la capacidad de combinar los constituyentes de los pensamientos es lo que constituye tener un concepto: “Así, alguien que piensa que *Juan es feliz* y que *Harry es feliz* ejercita en dos ocasiones la habilidad conceptual que llamamos poseer el concepto de “felicidad”. Evans analiza la posesión de conceptos en términos de las habilidades conceptuales que ejerce el sujeto. De manera que alguien que piensa que *Juan es feliz* y que *Harry es feliz* ejercita en dos ocasiones la misma “habilidad conceptual” que consiste en poseer el concepto FELIZ. Poseer un concepto consiste en esta capacidad (o habilidad conceptual, en términos de Evans) de combinarlo de manera irrestricta con otros conceptos. Cuando decimos que

⁴ Esto es así en el caso de las oraciones del lenguaje natural porque, según Evans (1982, 102), “un significado expresado por una oración estructurada también puede expresarse con una oración no estructurada”. Por ejemplo, una oración de una palabra introducida por estipulación para tener tal y cual significado.

un sujeto posee el concepto FELIZ, decimos que no hay ninguna restricción para que el sujeto piense *Juan es feliz, Harry es feliz, Susana es feliz, María es feliz*, etc. (teniendo en cuenta que posee los conceptos JUAN, HARRY, etc.).

Siguiendo a Ryle (1949), las habilidades, sean conceptuales o no, son propiedades disposicionales. Una propiedad disposicional no se encuentra en un estado particular, sino que es susceptible de encontrarse en un estado particular cuando se da determinada condición. Por ejemplo, que algo sea frágil no consiste en que está actualmente roto, sino que consiste en que, si fuese sometido a golpes o forcejeos, se rompería. En el caso específico del desempeño en ciertas habilidades conceptuales, aceptándose que las habilidades son disposiciones, no es un asunto de lo que efectivamente se hace, sino de lo que se es capaz de hacer con ese concepto (FODOR 1998, 19).

Las disposiciones son propiedades que pueden actualizarse bajo determinadas condiciones. En general, para cada disposición hay actualizaciones que parecen ser típicas. Así, la manera típica en que se actualiza la fragilidad de un objeto es rompiéndose cuando se lo somete a golpes. Pero quizá haya maneras menos típicas de actualización de la disposición de ser frágil. Someter un objeto a determinadas condiciones de calor que traigan como consecuencia una alteración en su estructura, tal vez sea una manera de actualizar su fragilidad. Lo que me importa resaltar es que las disposiciones pueden actualizarse de diferentes maneras gracias a su base categorial. La base categorial de una propiedad disposicional está constituida por las propiedades microestructurales que son causalmente responsables, en ciertas circunstancias, de las distintas manifestaciones de la disposición (FARA 2006).

Las habilidades conceptuales, de las que habla Evans, en tanto disposiciones tendrían una base categorial. La base categorial de las habilidades conceptuales es la que tiene la eficacia causal para producir las diferentes manifestaciones de estas habilidades. Evans (1982) admite que hay “algo en común” entre el pensamiento *Fa* y *Ga*, en el sentido de que cada vez que se ejercita *a* debe haber una causa común para esa habilidad. Con todo, Evans no hace un tratamiento de las habilidades conceptuales poniendo peso en su base categorial. Para poder decir algo más en relación a la base causal que subyace a las habilidades conceptuales habría que alejarse de la presentación que hace Evans del requisito de generalidad.

Tenemos que explicar de dónde surgen los poderes causales comunes que permiten pensar *Fa*, *Fb*, *Ga* y *Gb*. Se puede pensar que las habilidades de pensar *Fa*, *Fb*, *Ga* y *Gb* se dan gracias a las representaciones conceptuales que tienen el poder causal de generar estas distintas combinaciones. Estas representaciones tienen ese poder causal debido a sus propiedades sintácticas. Y las propiedades sintácticas/causales son propiedades del vehículo de las representaciones. Según esta

lectura, el requisito de generalidad es un requisito que satisfacen los vehículos de los pensamientos (BECK 2012). Esta es la lectura que adopta Fodor y Phyllyshyn (1988) del requisito. A diferencia de Evans, para quien poseer un pensamiento consiste en ejercer ciertas habilidades cognitivas estructuradas, para Fodor poseer un pensamiento consiste en la combinación de particulares mentales.

La caracterización filosófica de los conceptos a partir del requisito de generalidad es insuficiente si tenemos en cuenta la investigación empírica. Difícilmente puedan encontrarse referencias al requisito de generalidad en los estudios psicológicos sobre categorización y otras tareas cognitivas en las que intervienen conceptos. Como creo que una buena manera de abordar las representaciones conceptuales es atendiendo tanto a una especificación filosófica como psicológica en lo que sigue desarrollaré en qué sentido el requisito de sistematicidad se asocia a un aspecto fáctico de nuestra vida cognitiva. Será la propuesta de Jerry Fodor la que nos ayude a tal propósito. El requisito de generalidad expresa, en parte, la propiedad de la sistematicidad que Fodor (1975) atribuye a los pensamientos (BECK 2012; CAMP 2004). Si una mente puede pensar que *Juan ama a María*, también puede pensar que *María ama a Juan* (FODOR 1998, 139). Asimismo, si una mente puede pensar que *María es feliz* y *Juan es austero*, también puede pensar que *María es austera* y *Juan es feliz*. El requisito de generalidad hace referencia, en parte, al mismo fenómeno de la sistematicidad. Existe al menos una diferencia entre el requisito y el fenómeno de la sistematicidad. El primero es una propuesta conceptual acerca de cómo deben estar estructurados nuestros pensamientos. Evans (1982) entiende que es un principio *ideal* al cual los sistemas de pensamiento sólo se ajustan de manera aproximada. El segundo, en cambio, está planteado como un hecho empírico acerca de los pensamientos de las mentes humanas. Parece conceptualmente posible que haya mentes que aun pensando que *Juan ama a María*, no tengan la capacidad de pensar que *María ama a Juan*. Sin embargo, tener la habilidad de pensar el segundo pensamiento a partir del primero es “un rasgo de la vida mental de los humanos” (FODOR 1998, 49).

La explicación de la sistematicidad está dada, en parte, por una cuestión que atañe al vehículo de esos pensamientos, vehículo que Fodor entiende en términos del lenguaje del pensamiento.⁵ Así, los elementos mínimos del len-

⁵ La sistematicidad se explica fundamentalmente por la composicionalidad sintáctica. Esto no implica que no intervenga la composicionalidad semántica. La sistematicidad es una propiedad del pensamiento que se explica en función de la composicionalidad semántica y sintáctica. Por un lado, un sistema de representaciones es semánticamente composicional en caso de que las propiedades semánticas de la representación compleja estén totalmente determinadas por las propiedades semánticas de sus constituyentes. Por otro lado, un sistema de representaciones es sintácticamente composicional en caso de que las propiedades sintácticas de la representación compleja estén totalmente determinadas por las propiedades sintácticas de sus constituyentes. Ver Fodor (2008)

guaje del pensamiento (los conceptos) se recombinan según el requisito de generalidad. Evans, en cambio, cree que no hace falta comprometerse con la idea de que los pensamientos se vehiculan en lenguaje del pensamiento debido a que, como heredero de Wittgenstein, se aparta de la idea de que los conceptos son representaciones mentales.⁶ Desde la perspectiva fodoriana, en cambio, el requisito de generalidad hace alusión esencial a los vehículos representacionales del lenguaje del pensamiento. Creo que esta idea fodoriana es satisfactoria sólo parcialmente. En lo que sigue me ocuparé de los aspectos que considero correctos e incorrectos de la presentación fodoriana del requisito de generalidad.

El tratamiento que hace Fodor de la sistematicidad o requisito de generalidad parece mostrar la necesidad de vehículos lingüísticos. Siguiendo a Fodor (2007; 2008), considero que las representaciones mentales cuyos vehículos son perceptivos no satisfacen el requisito. Creo que una buena razón para sostener esto es la siguiente. Si se admite que el requisito de generalidad se refiere a la forma de ciertos pensamientos, entonces las combinaciones a las que alude el requisito se hacen en virtud de la forma lógica de ciertos pensamientos. En términos muy generales, los pensamientos tienen una forma lógica subyacente que es una representación formal de su estructura lógica. La forma lógica expresa, de alguna manera, la estructura composicional del pensamiento (e.g., *Fa*, *Fb*, *Ga*, *Gb*) y lo hace explicitando la contribución que hace cada constituyente a la totalidad del pensamiento. Un defensor de los vehículos perceptivos, ¿puede dar cuenta de la forma lógica de los pensamientos?

La noción de “forma lógica” ha sido desarrollada desde un punto de vista filosófico y lingüístico (CARRUTHERS 2003). Desde una perspectiva filosófica, la forma lógica de una oración es una construcción que contiene constantes lógicas, cuantificadores, variables y constantes de individuo. Desde una perspectiva lingüística, la forma lógica es una representación sintácticamente estructurada que es el resultado de los cálculos lingüísticos. La forma lógica junto con la forma fonética forman una descripción estructural que provee de instrucciones a los sistemas de actuación (CHOMSKY 1995). Ahora bien, independientemente de cómo entendamos la forma lógica, ésta no puede extraerse de las representaciones perceptivas.

La forma lógica puede extraerse de aquellas representaciones cuya descomposición es canónica. Por ejemplo, dadas las representaciones lingüísticas “María es dentista” y “Pedro es artista”, ambas se pueden descomponer en términos singulares y predicados. Siempre que hacemos el ejercicio de descom-

⁶ Davies (1998), es uno de los que ha argumentado que los neo-fregeanos, como Evans, están comprometidos con el lenguaje del pensamiento. Esto daría cuenta de la posibilidad de vincular las teorías de los conceptos concebidos como entidades abstractas con aquellas que asumen que son particulares mentales.

poner estas representaciones encontramos los mismos componentes, términos singulares y predicados, y a estos componentes que siempre son los mismos en la descomposición de estas representaciones se los llama constituyentes canónicos. Como estas representaciones tienen constituyentes canónicos, se combinan de manera canónica en virtud de estos constituyentes y no en virtud de ningún otro tipo de constituyente. Las combinaciones canónicas que surgen de estas representaciones serían “María es artista” y “Pedro es dentista”. Las representaciones de las cuales puede extraerse una forma lógica son las que tienen estas características, dado que la forma lógica hace explícita las relaciones entre los constituyentes canónicos de una representación (FODOR 2008).

Las imágenes mentales tienen características diferentes. Desarrollar este punto contribuirá a entender por qué las simulaciones no cumplirían con la condición de la sistematicidad. Pues, considérese que desde el enfoque enraizado la imaginación mental constituye uno de los mecanismos que opera haciendo uso de simulaciones (BARSALOU 2008). Una imagen mental (visual, motora o auditiva) es una representación consciente de un objeto ausente (BARSALOU 1999A). Tomando como ejemplo a las imágenes visuales, las mismas no pueden descomponerse canónicamente. Se pueden descomponer de diversas maneras y en cada descomposición surgen componentes diferentes. Siguiendo un ejemplo de Fodor (2008), la foto de una persona (una imagen gráfica que tiene las mismas propiedades de las imágenes visuales) puede recortarse en partes de muchas maneras, y en cada recorte surgirían componentes diferentes de la foto. La descomposición de una imagen gráfica el carácter no canónico de sus partes. Dado que en las diferentes descomposiciones de una misma imagen pueden surgir diferentes componentes, estos componentes no serían constituyentes canónicos sino que, en cambio, serían meras partes a las que subyacen simulaciones. Por supuesto que las partes de una imagen visual pueden combinarse gracias al procesamiento subyacente de simulaciones. Siguiendo el ejemplo de la foto, los pedazos de la foto recortada se combinan para representar la imagen total de una persona. Pero como estas combinaciones se dan entre meras partes y no entre constituyentes canónicos, entonces las simulaciones subyacentes se combinarían de manera no canónica. De representaciones con estas características no se podría extraer ninguna forma lógica, lo cual significa que las imágenes visuales, a las cuales subyacen las simulaciones, no satisfacen el requisito de generalidad, el cual alude a las combinaciones realizadas en virtud de la forma lógica. Es cierto que Barsalou (1999A) considera que las simulaciones se combinan de manera sistemática pero tal como lo afirman Adams y Campbell (1999) el autor no explica de qué manera podrían satisfacer esta condición para la conceptualidad.

Habiendo desalentado la idea de que el requisito de generalidad se explica postulando simulaciones, creo que la posibilidad que queda es la que pone

en evidencia Fodor (2007; 2008). Si los pensamientos como *Juan es feliz* tienen la forma lógica *Fa*, esto es posible porque este pensamiento tiene una descripción canónica de la cadena de símbolos. La descripción canónica de *Juan es feliz* es ((Juan)_{SN} ((es)_V (feliz)_{SP})_{SV})_O. Esta descripción determina cuáles son los constituyentes canónicos del pensamiento *Juan es feliz*, eliminando constituyentes no canónicos como *Juan es*. Asimismo, la descripción canónica establece relaciones de jerarquía entre los constituyentes canónicos de *Juan es feliz*. Cada uno de estos constituyentes contribuye de diferente manera a la totalidad del pensamiento. El constituyente *Juan*, en tanto núcleo nominal, contribuye de manera diferente que el constituyente *es*, en tanto núcleo verbal.

La idea es que esta descripción canónica es posible si el vehículo de ese pensamiento es lingüístico. Sólo en un medio representacional lingüístico se pueden establecer constituyentes canónicos y relaciones jerárquicas entre los constituyentes de los pensamientos. El vehículo lingüístico permite una sintaxis combinatoria de los pensamientos, según la cual los constituyentes de los pensamientos tienen relaciones estructurales entre sí. Asimismo, los pensamientos vehiculados por representaciones lingüísticas tienen una semántica composicional, según la cual sus constituyentes siempre hacen la misma contribución semántica en el pensamiento donde aparecen. La sintaxis combinatoria y la semántica composicional son características de los pensamientos que poseen constituyentes canónicos. Así, el requisito de generalidad requiere combinaciones que se hacen en base a la forma lógica de los pensamientos, la forma lógica de los pensamientos requiere una descripción canónica de sus componentes, la descripción canónica requiere que el vehículo representacional sea lingüístico, por lo tanto, el requisito de generalidad requiere que el vehículo representacional sea lingüístico.

Creo que es correcto considerar, junto con Fodor, que el requisito de generalidad nos compromete con un vehículo lingüístico. Sin embargo, considero incorrecto creer, como hace Fodor, que la conclusión anterior nos compromete de manera inmediata con la hipótesis del lenguaje del pensamiento. Un vehículo que tiene sintaxis combinatoria y semántica composicional es oracional. Así, también puede considerarse que los vehículos del lenguaje natural tienen algún tipo de sintaxis combinatoria y semántica composicional.⁷ De hecho,

⁷ Fodor (2001), asumiendo que el pragmatismo radical respecto del lenguaje natural es correcto, sostiene que el lenguaje natural no es composicional. Según esta nueva perspectiva, solamente el lenguaje del pensamiento sería genuinamente composicional. Sin embargo, creo que este giro no es legítimo. Siguiendo a Clapp (2010), considero que si se admiten los argumentos pragmáticos radicales en contra de la composicionalidad del lenguaje natural, entonces también debería rechazarse la composicionalidad del lenguaje del pensamiento. Si el lenguaje natural no es composicional, Fodor dispondría de pocas herramientas para sostener que el lenguaje del pensamiento sí lo es.

Fodor se inspira en las propiedades de los lenguajes naturales para caracterizar al lenguaje del pensamiento como teniendo sintaxis combinatoria y semántica composicional. En este sentido, creo que el requisito de generalidad exige que las representaciones tengan vehículos lingüísticos, y estos pueden ser o bien del lenguaje del pensamiento o bien del lenguaje natural.

Resumiendo, los conceptos son aquellas entidades que satisfacen de manera conjunta las dos condiciones que expuse a lo largo de este apartado. Cabe aclarar que la sistematicidad y la independencia del estímulo son propiedades conceptualmente independientes:

Una criatura [...] podría recombinar sistemáticamente sus representaciones aun cuando carezca de la libertad respecto del control del estímulo. A la inversa, podemos imaginar criaturas cuyos pensamientos fueran producidos por representaciones totalmente separadas e inestructuradas, en lugar de ser producidos por la interacción sistemática de sus constituyentes (representacionales), pero que pudiera tener todos sus pensamientos potenciales, cualquiera sea la circunstancia actual (CAMP 2009, 291)

Desde un punto de vista conceptual estas propiedades son independientes, pero “en la práctica están mucho más íntimamente relacionadas” (CAMP 2009, 291). Así, las representaciones conceptuales son las que se combinan sistemáticamente y son independientes del estímulo.

5. CONCLUSIÓN

En un intento de elucidar la noción de “representación conceptual” como un aporte al debate enraizado, adopté un criterio funcional para la caracterización de la misma. Los roles funcionales básicos de los conceptos serían los siguientes: (i) se pueden combinar sistemáticamente para formar nuevas estructuras representacionales y (ii) se emplean en una variedad de tareas cognitivas con independencia del estímulo. Estas condiciones no son arbitrarias. Por el contrario, son condiciones que suelen aparecer en las propuestas psicológicas y filosóficas que caracterizan al contenido conceptual.

De modo que, las representaciones que son independientes del estímulo pero no pueden combinarse de manera sistemática no son conceptuales. Por ejemplo, tomemos el caso del estado propioceptivo en el cual un sujeto, después de una operación, aún siente dolor en su pierna amputada. El sujeto está en un estado en el cual siente que su miembro todavía está conectado al cuerpo y sigue funcionando con el resto de éste cuando en realidad es un

miembro fantasma. Este estado representacional es independiente del estímulo porque se instancia sin que se requiera la presencia del estímulo proveniente del miembro en cuestión. En este caso la independencia del estímulo se da en tanto que la representación se activa aunque en el mundo no existe el estímulo pertinente. Pero sería difícil sostener que la representación que constituye este estado propioceptivo puede combinarse sistemáticamente con otras para dar origen a otros estados propioceptivos. Un estado propioceptivo posee la misma naturaleza que un estado perceptivo y, como hemos visto, los estados perceptivos no poseen constituyentes canónicos que puedan combinarse de manera sistemática.

Otro ejemplo es la conducta de los atáxicos ópticos. Los pacientes con ataxia óptica, aunque no tienen problemas perceptivos ni de orientación, no pueden alcanzar y tomar objetos presentados visualmente (ROSSETTI *et al.* 2003). Suele considerarse que la ataxia óptica es una patología de la coordinación visuomotora, la cual está involucrada en las acciones intencionales en las que el sujeto produce deliberadamente una secuencia de movimientos para alcanzar un objeto (ROSSETTI *et al.* 2003). Debido a este trastorno para apuntar hacia un *target* visual, se les vuelve difícil a los sujetos desenvolverse en tareas muy simples como prender un cigarrillo o tomar un vaso de agua. La ataxia óptica parece involucrar representaciones motoras independientes del estímulo. Aun cuando haya un estímulo distal presente, las representaciones motoras de los pacientes con ataxia óptica no están guiadas por ese estímulo. Pero estas representaciones no pueden combinarse sistemáticamente. La razón se debe a que los vehículos de las representaciones motoras no presentan la estructura de constituyentes canónicos que hemos visto.

Asimismo, las representaciones que pueden combinarse de manera sistemática y no pueden ser independientes del estímulo, tampoco serían conceptuales. Puede concebirse el requisito de generalidad de manera contrafáctica, es decir, según la restricción de la siguiente cláusula: “si ocurriera el estímulo relevante, entonces se combinarían las representaciones de la manera relevante” (CAMP 2009, 290). Esta cláusula restringe el dominio de combinaciones sistemáticas a las representaciones que un organismo activa en virtud del “estímulo relevante”. Camp (2009) considera que ciertos animales con cognición básica satisfacerían el requisito de generalidad bajo esta cláusula contrafáctica. Siguiendo un ejemplo presentado por Camp (2009, 280), el perro *D* puede representarse al perro *M* como amenazante y al perro *N* como amistoso así como también puede representarse al perro *M* como amistoso y al perro *N* como amenazante. Estas combinaciones representacionales estarían restringidas por la información del medio ambiente, en el sentido de que se dan debido a la conducta efectiva que evidencian estos animales. De modo que, por ejemplo,

si ocurriera que el perro *M* evidenciara una conducta amenazante, entonces el perro *D* lo representaría como amenazante. Lo importante a destacar es que este es un caso en el que se cumpliría el criterio de sistematicidad pero no se cumpliría el criterio de la independencia del estímulo para la conceptualidad. Así, según esta propuesta, las representaciones no conceptuales serían aquellas que no satisfacen en conjunto los criterios de independencia del estímulo y sistematicidad.

La idea que se asume en este artículo es que esta clarificación conceptual contribuiría a echar alguna luz sobre el debate metodológico enraizado en el cual se polemiza sobre la manera de integrar el enfoque enraizado con el clásico. El enfoque enraizado ofrece una explicación de las capacidades conceptuales en el cerebro apelando a las modalidades perceptivas, motoras, afectivas e introspectivas (BARSALOU 2016). Esto no significa que haya necesariamente una reducción de nuestras capacidades conceptuales a los mecanismos sensorio-motores. El objetivo del enfoque enraizado sería entender cómo las modalidades así como también el ambiente físico y social contribuyen a la cognición pero dejando un espacio para las explicaciones que apelan a representaciones simbólicas (BARSALOU 2007). Buscar si hay univocidad o no en la noción de “representación conceptual” en primer lugar contribuiría a definir si efectivamente existe un acervo conceptual que compartan ambos enfoques. Esto es fundamental para vislumbrar si hay un verdadero debate o simplemente un diálogo de sordos entre partes. Como se desarrolló, las simulaciones sólo parcialmente responden a los criterios de conceptualidad elucidados dado que son independientes del estímulo pero no se combinan de manera sistemática. Aunque se necesitarían más elementos de juicio, podría pensarse que la manera de integrar estos enfoques es considerando que las representaciones simbólicas, que responden paradigmáticamente a las condiciones de conceptualidad analizadas, se complementan en la arquitectura cognitiva con las representaciones enraizadas no conceptuales. Otra posible interpretación es considerar que las simulaciones son entidades neurales en las que se implementan las funciones propias de las arquitecturas cognitivas clásicas. Johnson (2018) entiende que si bien Barsalou presenta su teoría en términos representacionales, bien podría entenderse como una propuesta neuropsicológica. Con todo, aún quedan muchos aspectos abiertos en este debate metodológico enraizado.

REFERENCIAS

- ADAMS y CAMPBELL 1999, "Modality and Abstract Concepts", *Behavioral and brain sciences* 22: 577-660.
- BARSALOU, L. W. 1999A, "Perceptual Symbol System", *Behavioral and brain sciences* 22: 577-660.
- BARSALOU, L. W. 1999B, "Perceptions of Perceptual Symbols", *Behavioral and brain sciences* 22 (4).
- BARSALOU, L. W. 2008, "Grounded Cognition", *Annu. Rev. Psychol.* 59: 617-45.
- BARSALOU, L. W. 2010, "Grounded Cognition: Past, Present, and Future", *Topics in Cognitive Science* 2: 716-24.
- BARSALOU, L. W. 2016, "On Staying Grounded and Avoiding Quixotic Dead Ends", *Psychon Bull Rev* 23: 1122-42.
- BARSALOU, L. W. *et al.* 2007, "Cognition as Coordinated Non-cognition", *Cognitive Processing* 8: 79-91
- BARSALOU, L. W. 1999, "Perceptual Symbol Systems", *Behavioral and Brain Sciences* 22: 577-660.
- BECK, J. 2012, "The Generality Constraint and the Structure of Thought", *Mind* 121(483): 563-600.
- BERMÚDEZ, J. L. 2003, *Thinking Without Words*, Oxford, OUP.
- BRAINE, M. D. S.; REISER, B. J. y RUMAIN B. 1984, "Some empirical justification for a theory of natural propositional reasoning", G. H. BOWER (ed.) *Psychology of Learning and Motivation*, Orlando, Academic Press.
- CAMP, E. 2009, "Putting Thoughts to Work: Concepts, Systematicity, and Stimulus-Independence", *Philosophy and Phenomenological Research* 2: 275-311.
- CAMP, E. 2004, "The Generality Constraint: Nonsense and Categorical Restrictions", *Philosophical Quarterly* 54: 209-31.
- CARRUTHERS, P. 2003, "Thinking in Language? Evolution and a Modularist Possibility", P. CARRUTHERS y J. BOUCHER (ed.), *Language and Thought*, Cambridge, CUP.
- CHOMSKY, N. 1995, *El programa minimalista*, Madrid, Alianza.
- CLAPP, L. 2010, "Is even Thought Compositional?", *Philosophical Studies* 10.1007/s11098-010-9649-2, Online First.
- CONCHIGLIA, G.; DELLA ROCCA, G. y GROSSI, D. 2007, "On a Peculiar Environmental Dependency Syndrome in a Case with Frontal-Temporal Damage: Zelig-like Syndrome", *Neurocase* 13: 1-5.
- DAVIES, M. 1998, "Language, thought and the language of thought (Aunty's own argument revisited)", P. CARRUTHERS y J. BOUCHER (ed.) *Language and Thought*. Cambridge, CUP.

- DOVE, G. O. 2008, "Beyond Perceptual Symbols: a Call for Representational Pluralism", *Cognition* 110: 412-31.
- ELLIS, A. W. y YOUNG, A. W. 1992, *Neuropsicología cognitiva humana*, Barcelona, Masson.
- ESTES, W. K. 1994, *Classification and cognition*, Oxford, OUP.
- EVANS, G. 1982, *Varieties of Reference*, Oxford, OUP.
- FARA, M. 2006, "Dispositions", *Stanford Encyclopedia of Philosophy* (edición primavera de 2006), E. N. Zalta (ed.), <http://plato.stanford.edu/entries/dispositions/>.
- FODOR, J. 2008, *LOT 2: The Language of Thought Revisited*, Oxford, OUP.
- FODOR, J. 2007, "The Revange of the Given", B. P. McLAUGHLIN y J. COHEN (ed.), *Contemporary debates in philosophy of mind*, Oxford, Blackwell.
- FODOR, J. 2001, "Language, Thought and Compositionality", *Mind and Language* 16: 1-15.
- FODOR, J. 1998, *Conceptos. Donde la ciencia cognitiva se equivocó*, Barcelona, Gedisa.
- FODOR, J. 1975, *El lenguaje del pensamiento*, Madrid, Alianza.
- FODOR, J. y PYLYSHYN, Z. W. 1988, "Connectionism and Cognitive Architecture: a Critical Analysis", C. McDONALD y G. McDONALD (ed.) (1995), *Connectionism*, MA, Blackwell.
- FREGE, G. 1892, "On Sense and Reference", P. GEACH y M. BLACK (ed.), *Translations from the Philosophical Writings of Gottlob Frege*, Oxford, OUP (1966).
- GENTNER, D. 1983, "Structure-mapping: A theoretical framework for analogy", *Cognitive Science* 7: 155-70
- GOLDINGER, S. D.; PAPESH, M. H.; BARNHART, A. S.; HANSEN, W. A. y HOUT, M. C. 2016, "The poverty of embodied cognition", *Psychon Bull Rev* 23: 959-78.
- GREENE, J. D. 2009, "Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie", *Journal of Experimental Social Psychology* 45: 581-4.
- JOHNSON, M. 2018, "The Embodiment of Language", NEWEN, A.; DE BRUIN, L. y GALLAGHER, S. (ed.), *The Oxford Handbook of 4E Cognition*, Oxford, OUP.
- LAURENCE, S. y MARGOLIS, S. (ed.) 1999, *Concepts. Core Readings*, Cambridge, MA, MIT Press.
- LHERMITTE, M. D. 1986, "Human autonomy and the frontal lobes. Part II: Patient behavior in complex and social situations: The "environmental dependency syndrome", *Annals of Neurology* 19: 335-43.
- LOUWERSEN, M. 2012, "Symbol Indeterminacy in Symbolic and Embodied Cognition", *Topics in Cognitive Science* 3: 273-302.
- MACHERY, E. 2009, *Doing Without Concepts*, Nueva York, OUP.
- MAHON, B. Z. 2015, "What is embodied about cognition?", *Language, Cognition and Neuroscience* 30: 420-9.

- MARTIN, A. 2016, "GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain", *Psychon Bull Rev* 23: 979-90.
- MILLIKAN, R. 2000, *On Clear and Confused Ideas: An Essay about Substance Concepts*, Cambridge, CUP.
- MONTI, M.; LAWRENCE, M. y OSHERSON, D. 2009, "The Boundaries of Language and Thought in Deductive Inferences", *PNAS* 106, 125554-9.
- PECHER, D. 2013, "The Perceptual Representation of Mental Categories", D. REISBERG (ed.), *The Oxford Handbook of Cognitive Psychology*, Oxford, OUP.
- PICCININI, G. y SCOTT, S. 2006, "Splitting Concepts", *Philosophy of Science* 73: 390-409.
- PRINZ, J. 2004, *Gut Reactions. A Perceptual Theory of Emotions*, Oxford, OUP.
- PRINZ, J. 2002, *Furnishing the Mind. Concepts and their Perceptual Basis*, Cambridge, MA, MIT Press.
- REHDER, B. 2003, "A causal-model theory of conceptual representation and categorization", *Journal of Experimental Psychology Learning, Memory, and Cognition* 29: 1141-59.
- ROSH, E. 1999, "Principles of Categorization", LAURENCE, S. y MARGOLIS, E. (ed.), *Concepts. Core Readings*, Cambridge, MA, MIT Press.
- ROSH, E.; MERVIS, C. B.; GRAY, W. D.; JOHNSON, D. M. y BOYES-BRAEM, P. 1978, "Basic Objects in Natural Categories", *Cognitive Psychology* 8: 382-439.
- ROSSETTI, Y.; PISELLA, L. y VIGHETT, A. 2003, "Optic Ataxia Revisited", *Experimental Brain Research* 153: 171-9.
- RYLE, G. 1949, *El concepto de lo mental*, Buenos Aires, Paidós.
- SMITH, E. y MEDIN, D. L. 1981, *Categories and Concepts*, Cambridge, Harvard University Press.
- WEBER, A. y VOSGERAU, G. 2018, "Critical Note", NEWEN, A., DE BRUIN, L. y GALLAGHER, S. (ed.) *The Oxford Handbook of 4E Cognition*, Oxford, OUP.
- WEISKOPF, D. (ms.) "Concepts as a Functional Kind".
- ZWAAN, R. 2014, "Embodiment and Language Comprehension: Reframing the Discussion", *Trends in Cognitive Sciences* 18: 229-34.

BOSCO CORRALES TRILLO
Universidad Católica de Valencia "San Vicente Mártir"

(Not) Hosting Refugees in the EU: Philosophical Perspectives on the 2015-2016 Crisis

*La (no) acogida de personas refugiadas en la UE: perspectivas filosóficas
sobre la crisis de 2015-2016*

Recibido: 3/2/2021. Aceptado: 13/5/2021

Abstract: Were the EU's actions in terms of refugee hosting during the 2015-2016 crisis morally acceptable? This paper describes those actions and evaluates them from the points of view of philosophers who specialize in the field of asylum ethics, such as Joseph Carens, Matthew Gibney, or David Miller, among others. First, I clarify who, according to such literature, would be entitled to refugee or refugee-like protection. Secondly, I indicate to what those persons would be entitled. Thirdly, I explain why the EU fell short of its moral obligations towards those people, and finally, I advance a few suggestions for improvement.

Resumen: ¿Fue moralmente aceptable la actuación de la UE en cuanto a la acogida de personas refugiadas durante la crisis de 2015-2016? Este artículo describe dichas acciones y las evalúa desde el punto de vista de autores como Joseph Carens, Matthew Gibney o David Miller, entre otros. En primer lugar, aclaro qué personas tendrían derecho a protección internacional según los mencionados autores. En segundo lugar, indico a qué tendrían derecho aquellas personas. En tercer lugar, explico por qué la UE no estuvo a la altura de sus obligaciones morales hacia ellas y, por último, apunto algunas sugerencias de mejora.

Keywords: ethics, refugees, international protection, European Union, ideal and non-ideal theory.

Palabras clave: ética, personas refugiadas, protección internacional, Unión Europea, teoría ideal y no ideal.

INTRODUCTION

IN JUNE 2015, Nils Muiznieks, the Council of Europe's Commissioner for Human Rights, wrote a very critical op-ed article in *The New York Times* with the title "You are better than this, Europe" (MUIZNIEKS 2015). He considered that the response of most European countries to the refugee crisis had been shameful. Immigration management is certainly a complex and difficult issue, but we "must not use this difficulty as an excuse to trample on our obligations to protect those who flee wars and persecution", he said. During the following year, he repeatedly denounced what he considered to be outrageously immoral actions by the EU and its Member States (MUIZNIEKS 2016A; 2016B; 2016C).

I believe that Europe—as Mr. Muiznieks put it—trampled on its moral obligations towards refugees. However, as I try to make the case in the form of a philosophical research paper, the issue is so complex, the items to assess are so varied, and the opinions and arguments by the different authors so numerous and divergent, that the attempt at a rigorous and systematic ethical assessment of the EU's performance in the refugee crisis of 2015-2016 presents itself as an overwhelming task. This paper is just a first and modest—but thorough—approach to the issue, based on the theoretical proposals of some well-established philosophers, who specialize in the field of asylum ethics.

First, I shall offer some highlights of the EU's and its Member States' hosting actions during the crisis (section 1), in order to, then, explore to what extent they fulfilled their moral obligations, according to some accounts of moral responsibility to refugees in the recent, specialized philosophical literature. This I do in three steps: I shall first define who is entitled to international protection (section 2), secondly, I shall indicate to what they are entitled (section 3), and thirdly, I shall explain why, according to surveyed philosophical literature, EU States failed to fulfill their moral obligations to refugees in terms of hosting (section 4). Finally, I shall offer a brief conclusion (section 5) and advance a few suggestions on how EU States could or should have acted, based on the accounts of moral responsibility previously outlined; these suggestions are also meant as modest proposals to improve current refugee policies (section 6).

1. HOW DID THE EU AND ITS MEMBER STATES RESPOND TO THE REFUGEE CRISIS OF 2015-2016?

In this section, I shall indicate some of the most relevant actions of the EU and its Member States in terms of hosting refugees and contributing to fund protection elsewhere.

With regard to hosting refugees, EU institutions were not able to accomplish much, due to the very different stances of the Member States concerning asylum. After much protracted debate, the European Council managed to introduce relocation quotas to alleviate pressure on the EU Member States that served as entry points —mainly Greece and Italy—. First, the Council agreed on the relocation of 40,000 persons in need of international protection, from Italy and Greece to other EU Member States (COUNCIL OF THE EU 2015B). A few months later, due to the rise in arrivals of asylum seekers, a resolution was adopted to relocate additional 120,000 people (COUNCIL OF THE EU 2015C). Previously, the Council had agreed on resettling 22,504 persons in need of international protection from outside the EU (COUNCIL OF THE EU 2015A). However, most of those people ended up not being relocated, as we will see next.

EU Member States acted in very different ways, with regard to admitting asylum seekers and granting them international protection. In total, the whole of the EU countries granted asylum or other forms of international protection to over 900,000 people during 2015 and 2016 (EUROPEAN PARLIAMENT 2017). However, most of the effort was done by five or six countries, while many of the others dragged and lagged. Germany granted protection to approximately half of those 900,000¹ refugees in that two-year period; in addition, by the end of 2016, it had admitted a total of 587,346 asylum seekers with pending cases in its territory, thus becoming the champion of asylum during the crisis. The next top host countries in the EU during the 2015–2016 period were: Sweden, which granted protection to nearly 90,000 new refugees; France and Italy, each of them granting official protection approximately to 50,000 people; then, Greece, with roughly 40,000; Austria, over 30,000; the Netherlands, around 20,000; Denmark, 15,000; and Belgium, 12,000. Conversely, the main trend in most EU countries was very different. Some countries only granted asylum to a few hundred and some, to none. For instance, Ireland, Poland, and Croatia did not grant asylum to any new applicant. Latvia granted asylum to 149 people; Slovenia, to 162; Slovakia, to 190; Estonia, to 222; Lithuania, to 288; Portugal, to 494; and Czech Republic, to 644. The other EU States granted asylum to 1,000–7,000 applicants, but many of them were already there before the refugee crisis broke loose in 2015. Spain, for example, granted asylum to over 6,000 people, but most of them were already there before 2015. In fact, within the frame of the 2015 relocation quotas, Spain had committed to grant asylum to 17,337 refugees before September 2017, but it only granted it to 1,980. Similarly, many EU States did not meet their quotas and the EU relocation plan was a complete failure: out of the 160,000 refugees that EU Mem-

¹ The figures in this paragraph have been calculated comparing the UNHCR data from 2016 and 2014. See UNHCR (2017c, 6-9, Table I) and UNHCR (2015, 80-3, Table I).

ber States had agreed to relocate before September 2017, they only relocated 31,503 (EUROPEAN COMMISSION 2017B).

Concerning refugee protection elsewhere, the EU is the third top donor to the UNHCR worldwide, with donations made in the amount of 436 million USD in 2017, the first donor being the USA (with over 1.45 billion) and the second, Germany (with over 576 million) (DONORS n.d.).

Besides donating to the UNHCR, the EU humanitarian funding was remarkable. For instance, in 2016 alone, EU Civil Protection and Humanitarian Aid gave almost 2 billion for the support of forcibly displaced people and their host communities in 56 countries (Turkey, Greece, Syria, Iraq and South Sudan being the top five) (EUROPEAN COMMISSION n.d.). During 2016 and 2017, 1.4 billion euros were provided by the EU to fund humanitarian aid under the Facility for Refugees in Turkey (TURKEY n.d.). Since the beginning of the Syrian war and until 2017, 340 million euros had been given for the protection of refugees in Jordan (JORDAN n.d.) and 519 million for Lebanon (LEBANON n.d.). For the period 2015-2017, the EU humanitarian funding for both displaced Iraqis and Syrian refugees inside Iraq, totaled 349 million euros (IRAQ n.d.). From 2009 to 2017, 554.4 million euros had been dedicated to provide humanitarian aid in Pakistan for Afghan refugees, as well as for Pakistani people affected by the conflict in the northwest and by food insecurity (PAKISTAN n.d.). In 2017, the EU allocated 85 million euros to provide help for the more than one million South Sudanese refugees in Uganda, as well as Congolese, Burundian, and other refugees in that country (UGANDA n.d.). These are only some of the most significant EU actions in relation to protecting refugees elsewhere. We can safely affirm that the EU is very much committed to this line of protection.

In conclusion, we can say that the EU and some Member States contributed greatly to the protection of refugees. However, depending on how we define the EU's and its Member States' moral obligations toward refugees, we may conclude that most States—and even the EU in some aspects—failed to fulfil those obligations.

I shall next define who was entitled to protection by EU States and to what they were entitled, and then, assess whether those States lived up to their responsibilities toward them.

2. WHO WAS ENTITLED TO INTERNATIONAL PROTECTION IN THE EU?

The Universal Declaration of Human Rights states that “everyone has the right to seek and to enjoy in other countries asylum from persecution” (Art. 14, 1). It is true that, according to the Geneva Convention, a “refugee” is a forcibly

displaced person who is being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion.² In this sense, only a small number of forcibly displaced persons (FDPs) are refugees, strictly speaking, and have therefore a right to seek and enjoy asylum. However, there are other legal concepts that express the States' duty to host FDPs seeking protection.

In EU Member States, "when examining applications for international protection, the determining authority shall first determine whether the applicants qualify as refugees and, if not, determine whether the applicants are eligible for subsidiary protection" (DIRECTIVE 2013/32/ EU). Any FDPs who, if returned to their countries, would face real risk of serious harm, are eligible for subsidiary protection (DIRECTIVE 2011/95/ EU, art. 2, e).³ Serious harm includes, among other instances, "serious and individual threat to a civilian's life or person by reason of indiscriminate violence in situations of international or internal armed conflict" (DIRECTIVE 2011/95/EU, art. 15, c).

This status would apply to most of the, approximately, 6 million Syrian nationals forcibly displaced outside of Syria at the end of 2016, or to most of the millions of Afghan, Iraqi, South Sudanese citizens and a sadly very long list of other nationals forced to leave their countries due to violence and war. As a matter of fact, 50-61% of the more than 2.5 million asylum seekers that filed applications in EU countries during 2015-2016 were found eligible for international protection (EUROPEAN PARLIAMENT 2017). Several other millions were not allowed to arrive; had they reached an EU border, States would have had to process their cases and, given the situation in their countries of origin, it can be easily assumed that a similar percentage would have qualified for subsidiary protection (a status similar to asylum).

It is clear, from this legislation, that the EU and its Member States have very specific legal duties in terms of hosting and assisting refugees, as well as other kinds of FDPs. Consequently, even if one were to argue that article 14, 1 of the Universal Declaration of Human Rights does not necessarily imply for every country a duty to host, and even if most so called "refugees" would not qualify as such under the Geneva Convention, the EU has made surely ex-

² 1951 UN Convention Relating to the Status of Refugees, article 1 A(2): "Any person who: owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion, is outside the country of his nationality, and is unable to or, owing to such fear, is unwilling to avail himself of the protection of that country."

³ Directive 2011/95/EU: "A 'person eligible for subsidiary protection' means a third country national or a stateless person who does not qualify as a refugee but in respect of whom substantial grounds have been shown for believing that the person concerned, if returned to his or her country of origin, or in the case of a stateless person, to his or her country of former habitual residence, would face a real risk of suffering serious harm." This definition has been in force since 2004 (SEE COUNCIL DIRECTIVE 2004/83/EC).

plicit that its Member States do have the duty to host and protect *sensu stricto* refugees and many other people needing international protection of different kinds, when they apply for it.

There are other bodies of international law that either oblige or urge the EU to protect refugees and FDPs to different degrees,⁴ but the most recent, as well as the most comprehensive and demanding document with regard to refugee protection up to 2016 was the New York Declaration for Refugees and Migrants adopted by the UN General Assembly on September 19, 2016 (UNITED NATIONS 2016).⁵ Among the bold commitments contained in the document, I consider the pledge to “find new homes for all refugees identified by UNHCR as needing resettlement” (NEW YORK DECLARATION n.d.), to be the one that best exemplifies the moral obligation to render aid to refugees.

Not just convention refugees, but refugees in the broad sense of the term. The New York Declaration uses the term “refugee” in that sense, following the UNHCR definition of “refugee”. The UNHCR has enlarged the Geneva Convention definition of a “refugee”, to include what the EU legislation considers “persons eligible for subsidiary protection” (GLOSSARY n.d.).⁶

⁴ The European Convention on Human Rights prohibits the collective expulsion of aliens and prescribes certain procedural safeguards relating to the expulsion of individual aliens (European Convention on Human Rights, Protocol 4, art. 4 and Protocol 7, art. 1). This Convention is an international Treaty legally binding for all the members of the Council of Europe, which includes all EU countries. The EU itself is in the process of becoming a member too.

The Charter of Fundamental Rights of the European Union states that “the right to asylum shall be guaranteed with due respect for the rules of the Geneva Convention of 28 July 1951 and the Protocol of 31 January 1967 relating to the status of refugees and in accordance with the Treaty on European Union and the Treaty on the Functioning of the European Union” (art. 18). It adds that “collective expulsions are prohibited” (art. 19,1) and that “no one may be removed, expelled or extradited to a State where there is a serious risk that he or she would be subjected to the death penalty, torture or other inhuman or degrading treatment or punishment” (art. 19,2).

The Treaty on the Functioning of the European Union declares: “The Union shall develop a common policy on asylum, subsidiary protection and temporary protection with a view to offering appropriate status to any third-country national requiring international protection and ensuring compliance with the principle of non-refoulement” (art. 78, 1).

The UN General Assembly Resolution 69/167 of December 2014, stresses “the obligation of States to protect the human rights of migrants regardless of their migration status” and calls upon them to promote and effectively protect their rights and fundamental freedoms (“Resolution 69/167. Protection of Migrants,” UN General Assembly, December 18, 2014).

⁵ In December 2018 the UN agreed on a new deal known as “the global compact on refugees” (<https://www.unhcr.org/a-new-deal-for-refugees.html>), but it is beyond the scope of this paper, since we are trying to assess the EU’s response to the 2015-2016 refugee crisis.

⁶ The UNHCR has added the following to the Geneva definition: “(...) or [a person] who is outside his/her country of origin or habitual residence and is unable to return there because of serious and indiscriminate threats to life, physical integrity or freedom resulting from generalized violence or events seriously disturbing public order (OAU Convention and Cartagena Declaration)” (GLOSSARY n. d.).

This definition practically includes any person who, being forcibly displaced because of generalized violence, has crossed an international border. It virtually coincides with the definition of a “refugee” that authors like Matthew Gibney (2004 and 2018) or Joseph Carens (2013) use in their argumentations.

There is an ongoing academic discussion on whether the definition of “refugee” should be expanded or not.⁷ In a recent paper, Max Cherem argued that the refugee definition should not be expanded, because it would hurt the special case of refugees in the strict sense of the word and it would not necessarily benefit other kinds of forcibly displaced persons (CHEREM 2016, 187-96). He and others defending the same position —like M. Price, M. Lister, etc.— might perhaps be right in their contention against expanding the definition, but the fact is that the UNHCR has already expanded it and the UN Resolution 71/1 —signed by all EU States— is assuming such expansion. In any case, even if one were to argue that people fleeing generalized violence are not refugees, EU Member States would still have to act in compliance with the Council and Parliament Directives and apply the subsidiary protection status, which offers most of the benefits of refugee protection. Therefore, although the academic discussion is still open, I shall use the term refugee following the definitions of Gibney (2004 and 2018) or Carens (2013), because it essentially coincides with the UNHCR notion and with the EU definition of persons entitled to international protection in its Member States (going forward I shall use the term “refugee” in this extended sense, unless otherwise indicated).

3. TO WHAT ARE REFUGEES ENTITLED?

Some theorists, like Christopher H. Wellman, have argued that refugees only have a right to protection from their persecutors (WELLMAN and COLE 2011). This would only entail a safe haven where their life and physical integrity were not at risk. However, this contention is linked to both a narrow definition of refugees —exclusively as those fleeing persecution— and a rather partialist or nationalist understanding of States’ duties.

Conversely, most ethicists believe that refugees are entitled to the full range of human rights. At first, bare physical safety, food, and shelter can do in the short term, as immediate relieve. Then, however, civil and political as well as socio-economic rights must be progressively secured. This is the view not only of liberal cosmopolitans like Carens (2013) and others, but also of moder-

⁷ For a most recent and thorough discussion on the topic, see Cherem (2016) or Miller (2016, chapter 5). See also Gibney (2004 and 2018), Carens (2013), or Betts (2015), especially the text pertaining to notes 15-18 and the several studies referenced in those notes.

ate humanitarians like Gibney (2004), and even rather conservative communitarians like David Miller (2016). Also Martha Nussbaum, even considering cosmopolitanism to be a flawed theory, has recently insisted —quoting Grotius— that, “in a time of emergency our common claim, as human beings, to the wherewithal of life should dictate a liberal asylum policy” (NUSSBAUM 2019, 223). Furthermore, she considers that it is wrong to separate first and second-generation human rights, because without economic security, education, or health, it is not possible to exercise our civil and political liberties adequately.

Moreover, that refugees should be granted the full range of human rights is not only the conviction of philosophers, but it is contemplated in the law of any democratic State and very explicitly and particularly in the EU laws of international protection. According to Directive 2011/95/EU, arts. 20-35, both convention refugees and people eligible for subsidiary protection have a right to protection from refoulement, to receiving information, maintaining family unity, residence permits, travel documents, access to employment, access to education, social welfare, health care, and access to accommodation.

In addition, in many EU States, there is a third protection device called “authorization to stay for humanitarian reasons”, which entails full human rights protection. In other democratic countries, there are other forms of international protection, such as “temporary protected status” in the USA, or “persons in need of protection” in Canada, and so on. While not all these statuses include citizenship or permanent integration in the host society, they all guarantee civil liberties and socio-economic rights for the duration of the hosting time.

May these brief insights from different philosophical approaches and from democratic States’ legislation suffice to make clear that refugees are entitled to full human rights protection. However, this does not necessarily mean that such protection must be provided through asylum in rich countries. Funding and securing refugee protection in third countries can be an acceptable alternative, as we shall see.

4. DID EU STATES FULFIL THEIR MORAL OBLIGATIONS TOWARD REFUGEES?

The implication of the two previous sections is that, in 2015, EU countries were looking at several million people potentially arriving at its borders in just a few months and being entitled to enjoy protection in the EU if they applied for it. What should EU countries have done? Were they supposed to let them all come and apply for it? Where they morally obliged to take them

all in? Could they justifiably refuse to admit them? Under what criteria? Was externalizing borders and funding protection elsewhere a morally justifiable substitute for admitting refugees into the EU? I shall address these questions in the remainder of this paper.

Practically all theorists agree that States have some moral obligations to refugees (MILLER 2019). Most authors agree that refugee protection should be distributed equitably amongst States, so that each State is responsible for its fair share of refugee protection (GIBNEY 2004; CARENS 2013; MILLER 2019; and others). Some believe that developed countries should provide such protection mainly by hosting refugees (CARENS 2013; OWEN 2016A and 2016B; CORTINA 2017). Others consider protection elsewhere (securing refugee protection in a third country) to be an ethically acceptable substitute for hosting and integrating (BETTS and COLLIER 2017; GIBNEY 2004; MILLER 2016 and 2019; WELLMAN 2016). Michelle Foster, for instance, presents several acceptable protection elsewhere schemes (FOSTER 2007).

Did EU countries admit a fair amount of refugees during the 2015-2016 crisis? If not, did they fulfil their responsibilities to refugees by securing their human rights in third countries? I shall first discuss the hosting numbers and then turn to examine the protection elsewhere option.

It is not easy to ascertain any given country's fair share of refugee protection. GDP, total number of refugees in the world, and the host country population size, are commonly used as standards to measure a State's fair share of refugee resettlement (GIBNEY 2015, 9f.). GDP per capita measured in Purchase Power Parity would also be a very relevant item (OWEN 2016B), as well as causal responsibility (MILLER 2007, 86f.) among several other complex factors. In addition, if we want to measure not only the fair share of refugee protection by hosting, but also by outsourcing protection to third countries, the equation becomes even more complicated. The issue is intricate. However, for the sake of thinking the issue philosophically, without needing to be an expert in refugee demographics, perhaps I could try to indicate a very gross estimate of a plainly arithmetic distribution, to give ourselves a number from which we could start debating.

At the end of 2016, there were 17.2 million refugees under UNHCR mandate worldwide (UNHCR 2017B). If there are 194 countries in the world, each country should be responsible for the adequate protection of 88,659 refugees. By "adequate", I mean granting asylum or other forms of international protection in their own territory, or securing their protection in a third country that guarantees full human rights protection. This is a simple division and it is not indicative of how many refugees should a given country be really responsible for, but it seems to be a safe demand for rich democratic coun-

tries. If anything, the fair share figure should be higher for rich democratic countries and lower for poorer countries and for failed, refugee-producing, or crisis-ridden States. Yet, by the end of 2016, only eight EU States were hosting 88,000 refugees or more (UNHCR 2017B, 61-64, Annex Table I). Consequently, unless they were securing adequate protection elsewhere, the other EU countries' level of fulfilment of their moral obligations to refugees would have been questionable.

Admittedly, some authors and many governments will consider that securing adequate protection for 88,000 refugees is an excessive burden even for some EU States, depending on their economic growth, their public spending, public debt, unemployment rates, etc. For the sake of making prescriptions that could be acceptable for those governments, let's be extremely conservative and divide our simple arithmetic distribution by ten: 8,800. Again, according to the numbers provided by the UNHCR (2017B), only fourteen EU countries were hosting 8,800 refugees or more. I believe that, even being conservative, if they were not hosting a tenth of the arithmetic distribution and unless they were providing adequate protection elsewhere, it is safe to argue that most EU countries were probably not doing their fair share.

Before the European refugee crisis, I would have agreed with Carens, who did not see "how any democratic state in Europe or North America today could make the case that it has taken in so many refugees that it is now morally entitled to turn real refugees away" (CARENS 2013, 220). At the height of 2016, his statement needed further qualification. Germany—with 669,482 refugees and 587,346 pending asylum cases—had possibly reached and surpassed its fair share. France (approx. 370,000) or Sweden (313,000) could also make that case. Perhaps Italy, Austria, Greece, the Netherlands, and the UK could argue that they were doing their part too. However, Carens' evaluation would still apply to, at least, fourteen EU States, unless—again—they were effectively and adequately providing human rights protection in a third country. I shall discuss that possibility below, but first, we shall take a brief look at the world's refugee population.

If we look at the global picture, the unfair distribution of refugees looks much starker. At the end of 2016, after the greatest rise in asylum seekers worldwide, 84% of the refugees were being hosted by developing and least developed countries, while developed and rich countries were refusing to take in any more refugees. For instance, the EU, with approximately 10 times the GDP of Turkey, was hosting roughly the same number of people; with 20 times the GDP of Pakistan, the EU was hosting only twice as many people; with 12.5 times the GDP of Iran, it was hosting only three times as many peo-

ple; and with 335 times the GDP of Lebanon, it was hosting only three times as many people.⁸ The meager asylum figures in most European countries did not seem to justify a moral entitlement to turn refugees away.

Were EU States delivering their fair share of responsibilities by providing protection elsewhere? There is a number of reasons suggesting that protection in third countries could be a morally acceptable or even recommendable option. Some EU countries could lack absorptive capacity due to population density, inability to create enough jobs, economic crisis, potential political backlash, or public hostility. Another reason is that the cost of determining who is entitled to protection and providing such protection in rich countries is much higher than doing so in poor ones. Consequently, rich countries could afford providing protection for many more refugees if they outsourced their determination and protection mechanisms to developing ones. There are several other reasons in favor and against the moral rightness of providing refugee protection in third countries. In fact, there is an extensive debate on the matter (BETTS and COLLIER 2017; GIBNEY 2004, 215-20; CARENS 2013, 214-5; MILLER 2016, 85-93; OWEN 2016A; WELLMAN 2016).

Although the debate is completely open, I shall here assume —for the sake of the argument and for the sake of making prescriptions that real EU governments would feel compelled to follow— that there is nothing wrong per se with liberal democracies outsourcing (at least part of) their fair share of refugee protection. Even in that case, however, we would still have to introduce at least two provisions to make protection elsewhere schemes morally sound. First, they should be based on bilateral or multilateral agreements (CHEREM 2016; SANDELIND 2017); it would not be fair to just assume that any of the current host countries in the developing world must keep the refugees without its deliberate and explicit consent. Second, those schemes should guarantee full integration opportunities for refugees in the host society (SANDELIND 2017). Refugees are not only entitled to basic rights protection, like safety, food, or shelter. This may be enough in the short term, but in the long run, they are entitled to the full range of human rights, just

⁸ On the one hand, Turkey, with a GDP of 2 trillion international \$, was hosting 2.9 million refugees. Pakistan, with a GDP of 1 trillion, was hosting 1.4 million refugees. Lebanon, with 85.9 billion, was hosting 1 million. Iran, with 1.6 trillion, was hosting 979,400. On the other hand, the EU Member States, with 20.3 trillion, were hosting less than 1.9 million refugees and 1.1 million asylum seekers with pending cases. The GDP figures are presented in PPP (purchasing power parity) to compensate for the dissimilar cost of living in the different countries. See “GDP, PPP (current international \$),” World Bank, accessed June 4, 2018, https://data.worldbank.org/indicator/NY.GDP.MKTP.PP.CD?year_high_desc=true. The number of refugees hosted in each country are taken from UNHCR (2017c, 6-9, Table I)

like any other member of the society they are living in, which would ideally—but not necessarily—include permanent residency or even citizenship (CARENS 2013; GIBNEY 2004; OWEN 2016A and 2019). Therefore, without guaranteeing full integration opportunities—education, jobs, career, political participation, some kind of long-term residency, etc.—, the funding State would still be failing to fulfil its moral obligations towards refugees.

Betts and Collier (2017) consider that the existing refugee protection models in developing countries are dysfunctional and undermine the autonomy and dignity of refugees. Sandelind (2017, 19) has argued that virtually none of the protection elsewhere schemes of the EU and its Member States qualify as acceptable under these two provisions. Either they are not based on bilateral agreements or they do not guarantee integration (often, they do not even guarantee basic rights). Donating to the UNHCR does not guarantee refugee integration or fundamental rights protection. In fact, most of the time, such integration never happens, but refugees end up living in protracted situations of encampment without barely any prospects of being able to lead a dignified life. Funding refugee protection in third countries through humanitarian aid does not guarantee integration either and most of the time, is not based on bilateral agreements, but on the *de facto* assumption that refugees are staying in Lebanon or Jordan, for instance. The deal with Turkey would be an exception, since it was based on a bilateral agreement, but the scheme did not guarantee full protection of human rights and integration.

In fact, the 2016 EU-Turkey agreement has been considered by many as “shameful” (BULLEY 2017; GOGOU 2017). Strategically, it was successful: it reduced drastically the arrivals of migrants from Turkey to Europe and it can be said to be one of the most effective responses to the high rise in refugee flow. Ethically, however, it could be considered a low point in EU history of refugee protection and even a shame.

It was not shameful to ask Turkey to prevent departures from their coast by fighting the abominable human-smuggling mafias. Fighting crime was Turkey’s duty all along, which begs the question of why the need for an agreement with the EU, why Turkey was not fulfilling its duty, and whether it was neglecting it purposely. Not that Turkey had the duty to stop people from leaving the country, but most of those leaving by sea, were doing so with the help of unscrupulous mafias, which exploited their weakness, obtained huge illegal gains by extorting their money out of them, and finally let them on their own to a likely death by drowning in the sea. Again, the shameful part was not, as I said, asking Turkey to fulfil its duty and stop the criminal smuggling business. What was shameful was the fact that a union of 28 countries, with circa 500 million inhabitants, a GDP of 20 trillion,

and a GDP per capita of almost 40,000,⁹ refused to host more than 3 million refugees, while it asked Turkey to take care of the same amount of people: a country with 80 million inhabitants, a GDP of 2 trillion, a GDP per capita of 25,000, and an infant mortality rate of 11 (as opposed to the EU's 3),¹⁰ in addition to many other problematic issues.

It was not shameful to try and effectively reduce the migrants flow by sea, consequently preventing thousands of deaths. What was shameful was that the EU did not give the refugees any alternative option to apply for asylum, but it closed the door to hundreds of thousands of people in clear need of international protection, declaring that they would be returned if caught, which made the deal with Turkey not only ethically reprehensible, but also legally dubious. For these and other reasons, the EU-Turkey agreement continued to be strongly criticized by NGOs and migration policy experts, "who rightly questioned Turkey's status as a safe country of return, as well as a number of other issues of legality" (ARDITTI 2016). Also the Human Rights Commissioner of the Council of Europe repeatedly criticized a deal that he considered to elicit serious concerns from a human rights perspective and to be both morally and legally highly questionable (MUIZNIEKS 2016A; 2016B; 2016C).

It would have been ethically correct, if the deal had guaranteed human rights protection and provided long-term integration policies, but it only provided 6 billion euros for refugee protection, which, given the amount of refugees it required Turkey to take in, equaled to less than 2,000 euros per refugee. Such funding could barely cover the cost of short-term basic rights protection, let alone long-term full human rights protection.

For all these reasons, I am much inclined to concur with Cortina's assessment that the EU deal with Turkey was the result of "political panicking", rather than of well-pondered democratic debate (CORTINA and TORREBLANCA 2016).

In addition to asking Turkey to stem the flow of asylum seekers, other policies and measures were put in place to prevent refugees from reaching EU borders, without providing acceptable protection elsewhere as a remedial device. EU Member States implemented systematic, coercive, and prolonged detention of migrants and asylum seekers in contrast with international human rights law (MAJCHER 2017). The EU arranged a militarized naval blockade coordinated by its border agency Frontex, carried out operations to destroy smugglers' boats and return migrants to points of origin, and exercised pressure on North African

⁹ The figures are presented in PPP and measured in international \$. See "GDP per capita, PPP (current international \$)," World Bank, accessed June 4, 2018, <https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD>.

¹⁰ "Mortality rate, infant (per 1,000 live births)," World Bank, accessed June 5, 2018, <https://data.worldbank.org/indicator/SP.DYN.IMRT.IN>.

governments to create holding camps along the southern Mediterranean coast (HSU 2017, 8-9 and footnotes 16-19), without ensuring human rights protection for refugees in an adequate manner.

A Migration Partnership Framework was set up in June 2016 to support African countries of origin and transit with substantial funding and technical assistance, in order to help them stop irregular migration directed to the EU (EUROPEAN COMMISSION 2016). It was presented as a “new approach to better manage migration” with the objective “of saving lives” and yet, it was strongly condemned in a joint statement of 131 NGOs as being chiefly an attempt to stop migration from reaching the EU (Joint NGO statement). Although the EU communiqué explicitly declares to aim at “stemming the irregular flows while offering legal migration channels, including increased resettlement efforts”, such channels and efforts have been rather small. Even after having committed to finding “new homes for all refugees identified by UNHCR as needing resettlement” (NEW YORK DECLARATION n.d.) (UNITED NATIONS 2016, §78), out of the 1.2 million identified by UNHCR for 2018 (UNHCR 2017A), the European Commission recommended the resettlement of 50.000 until October 2019 (EUROPEAN COMMISSION 2017A). It is certainly an “increased” effort if compared with the less than symbolic 22,504 refugees the EU Member States agreed on resettling in 2015 (COUNCIL OF THE EU 2015A), but it is very far from enough, considering the real needs, the EU countries’ ability to host and their commitment to the New York Declaration.

These are some of the instances that show very clearly that the EU and most Member States were mainly working on strengthening the walls of a “Fortress-Europe” and externalizing borders in order to keep migrants and refugees out of the Schengen area, without ensuring ethically acceptable refugee protection elsewhere. The case for the EU and many of its Member States falling short of their obligations to refugees in terms of securing international protection seems to be rather compelling.

5. CONCLUSION

It must be acknowledged that both the EU and some Member States have contributed greatly to the protection of refugees and continue to do so. However, given the unprecedented rise in asylum seekers, in light of the comparison between the meager hosting numbers of the EU and those of host countries in the developing world, and considering the dereliction of adequate human rights protection elsewhere, we could say that many EU States and the EU itself failed to fulfil their moral obligations towards refugees in some

respects. Not only according to ideal approaches in normative theory, but also according to the very values on which the EU stands. The Treaty on the European Union (art. 2) establishes a very precise ethos as foundation for the EU:

The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail

In fact, the EU is often defined and it has often defined itself as a normative or ethical actor (BULLEY 2017, 53). Perhaps, one could argue that such an ethos is meant in a communitarian sense —along the lines of Walzer’s (1983) *Spheres of Justice*— and that the EU was always intended to protect its own, rather than others, with more responsibility for its people and Member States in an internal form of solidarity. However, the TEU itself declares that one of the Union’s central aims is to promote these values both at home and in “its relations with the wider world” by contributing to “peace, security”, and other social goods, including “solidarity and mutual respect among peoples” (art. 3.5). This is why, although the EU is meant as an area of freedom, security and justice primarily for EU citizens, the Tampere European Council recognized that “it would be in contradiction with Europe’s traditions to deny such freedom to those whose circumstances lead them justifiably to seek access to our territory” (TAMPERE EUROPEAN COUNCIL 1999).

Consequently, we could very well say, that the EU understands itself as guarantor of a kind of cosmopolitan ethics, because it is founded on the respect to human dignity and to the human rights of all persons and not only of its own citizens. Therefore, failing to admit a certain number of refugees or to secure their human rights protection elsewhere, would be a blatant contradiction with the value of human dignity and with the EU foundational values. Thus, some of the EU’s and its Member States’ actions during the refugee crisis of 2015-2016 could be considered unjust and an expression of “organized hypocrisy” (GIBNEY 2004, 229).

However, it would be utterly unrealistic to expect EU States to comply with the demands of an ideal notion of justice. As Cortina puts it, “the task of practical reason is not only to state what should be, but also to take flesh in the institutions, transforming them from the inside” (2010).¹¹ In this sense, the question is, what moral suggestions could realistically be expected to “take flesh” in the institutions of the EU and its Member States?

¹¹ The translation is mine.

6. SUGGESTIONS: WHAT SHOULD THEY HAVE DONE?

I present the following final remarks as merely brief sketches of some of the most relevant suggestions that can be drawn from the previous pages. They are by no means intended as an exhaustive list or a full explanation of proposals.

First, all EU countries should at least have honored their commitments in the frame of the EU relocation and resettlement plan by meeting their quotas and by doing so within the agreed deadlines. By failing to do that, their actions and omissions were utterly unethical.

Secondly, some EU countries should have admitted more refugees or should have made sure that a number of refugees enjoyed full human rights protection elsewhere—including long-term integration or a suitable substitute for it—. It is true that there is no international burden-sharing scheme to deal with refugee protection and it is also true that establishing fair share quotas for every country is an extremely complex issue (See MILLER 2019, 86-93 or GIBNEY 2004, 215-22). However, I believe the case can be made that, in the middle of the second greatest refugee crisis in history, granting asylum to zero refugees (Ireland, Poland, Croatia) would be considered unethical for any EU country, as probably would be hosting less than three hundred (Latvia, Slovenia, Slovakia, Estonia, Lithuania). That is, unless they effectively secured adequate protection elsewhere.

Admittedly, the magnitude of the refugee flow to Europe was alarming for EU governments and, in a non-ideal world, they were justified to fear severe consequences and long-lasting effects for their countries in terms of economic cost, political backlash, social unrest, etc. But even conceding that, the extremely restrictive and exclusive admission policies of some States should at least have been compensated by an effort to secure adequate protection in third countries.

Although in terms of ideal theory, I am inclined to agree that rich democratic countries should take up the slack and take in more refugees than their fair share would strictly demand (OWEN 2016), in the real-world scenario of the 2015 refugee crisis, I feel forced to agree with Miller (2019) that EU countries had no duty of justice to take up the slack, but merely to take in or otherwise effectively protect their fair share of refugees. Even Peter Singer, one of the most inclusive theorists, who in 2010 was still maintaining that affluent States should take in as many refugees as possible because the more fundamental needs of poor migrants were more important than the less fundamental ones of citizens (SINGER 2010), in 2015 proposed sending asylum claimants to refugee camps. He recognized that “it may not be the best solution, but it may be the most workable” (SINGER 2015).

The previous concessions, however, do not mean that European countries were morally justified to not comply with their obligations to refugees, by refusing to admit a minimum number of refugees and by failing to provide acceptable protection elsewhere. Singer's (2015) proposal was dependent upon two requirements: refugee camps in third countries must guarantee full human rights protection and such protection had to be paid by the affluent countries.

Massive refugee camps within the EU could have proven difficult to manage. If set up as open facilities, governments would have been looking at hundreds of thousands of asylum seekers potentially roaming the EU trying to settle down irregularly. On the other hand, if conceived of as detention camps, they would have been ethically most questionable from a liberal democratic perspective. Processing camps outside the EU, where asylum seekers are protected from persecution, their basic needs are covered, and they can undergo an effective and expedite eligibility process would perhaps have been best suited for an emergency situation like the 2015-2016 crisis.

Yet, in order for this suggestion to be ethically acceptable, processing camps could only be a short-term device. To avoid protracted encampment, resettlement should be a strict duty, like non-refoulement is (CARENS 2013). At the same time, fair share quotas of international protection should be established for every country, unless the international refugee system is to remain an "organized hypocrisy" (GIBNEY 2004). Such quotas should be calculated conservatively, so that they would only oblige States to take in a minimum of refugees, and with mandatory deadlines, so that the resettlement process flows and the encampment period does not lag, thus eroding refugee human rights protection.

A less inclusive consequence of this kind of policy is that it would oblige States to put in place, alongside effective determination processes, operational repatriation mechanisms for those who would be found not eligible. Again, this is not ideal. Ideally, I would side with unrestricted freedom of movement (CARENS 1992) and unconditional universal hospitality (CORTINA 2015), but even the most inclusive ideal theorists, when it comes to make prescriptions aimed at having an actual effect on real governments, accept restriction of movement and limits to hospitality (CARENS 2013; SINGER 2015).

Thirdly, the EU agreements with Turkey and with North African countries should have aimed more effectively at guaranteeing human rights protection for the millions of asylum seekers that were prevented from accessing EU territory by those same agreements.

Fourthly, EU States should have welcomed and even encouraged citizen involvement in resettling and integrating refugees. One of the main objections to increasing refugee protection efforts is that States have a greater duty with

regard to the interests of their own citizens than towards those of aliens. Yet, States' obligations to their citizens should work both ways and not only in the sense of restricting entry and downsizing refugee protection. Many EU citizens were interested and even eager to help refugees. In fact, civil society organizations played a major role in saving lives in the Mediterranean. The Spanish NGO Proactiva Open Arms, for instance, rescued 59,247 people from September 2015 to May 2018 (PROACTIVA OPEN ARMS n.d.).

In addition, other organizations worked actively and effectively to help resettling and integrating refugees in EU countries. For instance, the "Humanitarian Corridors" set up by the St. Egidio Community, together with the Waldensian and Methodist Churches, made possible the resettlement of thousands of refugees from Lebanon and Jordan to Italy, France, Belgium, and Andorra. The whole scheme was managed in compliance with EU immigration security measures and to no cost to any government. It also provided integration guarantees for refugees through their introduction into a wide social network, provision of health care, education, and access to the labor market (SANT'EGIDIO 2019). Unfortunately, most EU governments did not allow this kind of initiatives to take place in their countries. I believe that liberal democratic States should work in conjunction with active citizens with regard to providing refugee protection through search and rescue operations, first asylum, resettlement and integration schemes, or advocacy campaigns. States should not force their citizens to accept more refugees than justice demands, but neither should they impede that their citizens act effectively in favor of refugees.

Many other considerations could be made with regards to the EU response to the refugee crisis in 2015-2016. The obligations of other countries, the obligations of the UNHCR and other international organizations, the duties of refugees, and other issues could and should be taken into consideration. However, because of the complexity of the issue, I have decided to limit the scope of this paper to an ethical assessment of the EU's and its Member States' fulfilment of their moral obligations to host refugees during that crisis. On the one hand, such assessment has allowed us to conclude that some of their actions were ethically questionable or even unjust. On the other hand, it has permitted us to offer a limited number of suggestions on how they could have done better and could do better in the future.

BIBLIOGRAPHY

- ARDITTIS, S. 2016, "Is the EU Gradually Renouncing its Fundamental Values in Order to Better Protect its External Borders?", *Ethics and International Affairs*. Retrieved from <https://www.ethicsandinternationalaffairs.org/2016/eu-gradually-renouncing-fundamental-values-order-better-protect-external-borders> [accessed November 9, 2018].
- BARBULESCU, R. 2016, "Still a Beacon of Human Rights? Considerations on the EU Response to the Refugee Crisis in the Mediterranean", *Mediterranean Politics*. <http://dx.doi.org/10.1080/13629395.2016.1194546>
- BETTS, A. 2015, "The Normative Terrain of the Global Refugee Regime", *Ethics and International Affairs* 29(4). Retrieved from <https://www.ethicsandinternationalaffairs.org/2015/the-normative-terrain-of-the-global-refugee-regime/> [accessed May 3, 2018].
- BETTS, A.; COLLIER, P. 2017, *Refuge: Transforming a broken refugee system*, St. Ives: Penguin Book.
- BULLEY, D. 2017, "Shame on EU? Europe, RtoP, and the Politics of Refugee Protection", *Ethics and International Affairs* 31(1): 51-70. <https://doi.org/10.1017/S0892679416000666>
- CARENS, J. 1992, "Migration and morality: a liberal egalitarian perspective", B. BARRY and R. GOODIN (ed.), *Free movement: ethical issues in the transnational migration of people and money*, University of Pennsylvania Press, 25-47.
- CARENS, J. 2013, *The ethics of immigration*, Oxford: Oxford University Press.
- CHEREM, M. 2016, "Refugee Rights: Against Expanding the Definition of a 'Refugee' and Unilateral Protection Elsewhere", *Journal of Political Philosophy* 24(2): 183-205. doi: 10.1111/jopp.12071
- CORTINA, A. 2010, *Justicia cordial*, Madrid: Trotta.
- CORTINA, A. 2015, "Hospitalidad cosmopolita", *El País*, December 5.
- CORTINA, A.; TORREBLANCA, I. 2016, "Decálogo para la crisis de los refugiados", *El País*, March 10.
- CORTINA, A. 2017, *Aporofobia, el rechazo al pobre. Un desafío para la democracia*, Barcelona: Paidós.
- COUNCIL OF THE EU 2015A, "Justice and Home Affairs Council, 20/07/2015." Retrieved from <http://www.consilium.europa.eu/en/meetings/jha/2015/07/20/> [accessed May 28, 2018].
- COUNCIL OF THE EU 2015B, "Resolution of the Representatives of the Governments of the Member States meeting within the Council on relocating from Greece and Italy 40 000 persons in clear need of international protection." Retrieved from <http://data.consilium.europa.eu/doc/document/ST-11131-2015-INIT/en/pdf> [accessed June 2, 2018].

- COUNCIL OF THE EU 2015C, “Justice and Home Affairs Council, 22/09/2015.” Retrieved from <http://www.consilium.europa.eu/en/meetings/jha/2015/09/22> [accessed June 2, 2018].
- DONORS n.d., “UNHCR, The UN Refugee Agency.” Retrieved from <http://www.unhcr.org/donors.html> [accessed May 30, 2018].
- EUROPEAN COMMISSION 2016, “Communication from the Commission to the European Parliament, the European Council, the Council and the European Investment Bank on establishing a new Partnership Framework with third countries under the European Agenda on Migration”, June 7. Retrieved from http://ec.europa.eu/dgs/home-affairs/what-we-do/policies/european-agenda-migration/proposal-implementation-package/docs/20160607/communication_external_aspects_eam_towards_new_migration_ompact_en.pdf [accessed July 24, 2018].
- EUROPEAN COMMISSION 2017A, “Commission Recommendation of 27.9.2017 on enhancing legal pathways for persons in need of international protection.” Retrieved from https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/policies/european-agenda-migration/20170927_recommendation_on_enhancing_legal_pathways_for_persons_in_need_of_international_protection_en.pdf [accessed January 11, 2018].
- EUROPEAN COMMISSION 2017B, “Report from the Commission to the European Parliament, the European Council and the Council”, Annex 6.
- EUROPEAN COMMISSION n.d., “Forced displacement: refugees, asylum-seekers and internally displaced people (IDPs).” Retrieved from http://ec.europa.eu/echo/what-we-do/humanitarian-aid/refugees-and-internally-displaced-persons_en [accessed June 26, 2018].
- EUROPEAN PARLIAMENT 2017, *EU migrant crisis: facts and figures*, June 30. Retrieved from <http://www.europarl.europa.eu/news/en/headlines/society/20170629STO78630/eu-migrant-crisis-facts-and-figures> [accessed June 12, 2018].
- FOSTER, M. 2007, “Protection elsewhere: the legal implications of requiring refugees to seek protection in another state”, *Michigan Journal of International Law* 28: 223-86.
- GIBNEY, M. 2004, *The Ethics and Politics of Asylum. Liberal Democracy and the Response to Refugees*, Cambridge: Cambridge University Press.
- GIBNEY, M. 2015, “Refugees and Justice between States”, *European Journal of Political Theory*, online early: 1-17.
- GIBNEY, M. 2018, “The ethics of refugees”, *Philosophy Compass* 13(10): 1-9. <https://doi.org/10.1111/phc3.12521>
- GLOSSARY n.d., “UNHCR Global Focus.” Retrieved from <http://reporting.unhcr.org/glossary/r> [accessed February 12, 2019].

- GOGOU, K. 2017, "The EU-Turkey deal: Europe's year of shame." Retrieved from <https://www.amnesty.org/en/latest/news/2017/03/the-eu-turkey-deal-europes-year-of-shame/> [accessed March 23, 2018].
- HSU, R. 2017, "Global Europe: Rethinking migration, democracy, and what sustains Europe at home at abroad." Retrieved from: <http://recursos.ufv.es/docs/HSU-Roland-eng.pdf> [accessed February 17, 2019].
- IRAQ n.d., "European Civil Protection and Humanitarian Aid Operations of the European Commission website." Retrieved from http://ec.europa.eu/echo/where/middle-east/iraq_en [accessed May 5, 2018].
- JOINT NGO STATEMENT ahead of the European Council of 28-29 June 2016. NGOs strongly condemn new EU policies to contain migration, s.f., Oxfam website. Retrieved from https://www.oxfam.org/sites/www.oxfam.org/files/file_attachments/jointstatementeuimmigrationresponse.pdf [accessed June 6, 2018].
- JORDAN n.d., "European Civil Protection and Humanitarian Aid Operations", European Commission website. Retrieved from http://ec.europa.eu/echo/where/middle-east/jordan_en [accessed May 5, 2018].
- LEBANON n.d., "European Civil Protection and Humanitarian Aid Operations", European Commission website. Retrieved from http://ec.europa.eu/echo/where/middle-east/lebanon_en [accessed May 5, 2018].
- MAJCHER, I. 2017, "Border Securitization and Containment vs. Fundamental Rights: The European Union's 'Refugee Crisis'", *Georgetown Journal of International Affairs*. Retrieved from <https://www.georgetownjournalofinternationalaffairs.org/online-edition/border-securitization-and-containment-vs-fundamental-rights-the-european-unions-refugee-crisis>.
- MILLER, D. 2007, *National Responsibility and Global Justice*, Oxford: Oxford University Press.
- MILLER, D. 2016, *Strangers in our midst*, Cambridge, MA: Harvard University Press.
- MILLER, D. 2019, "Our responsibilities to refugees", M. HOESCH; L. LAUBE (ed.), *2018 ZIF Workshop "Studying Migration Policies at the Interface between Empirical Research and Normative Analysis"*, Bielefeld: ULB Münster, 37-49. doi:10.17879/95189441435
- MUIZNIEKS, N. 2015, "You're better than this, Europe", *The New York Times*, June 29. Retrieved from https://www.ny-times.com/2015/06/29/opinion/youre-better-than-this-eu-rop.html?_r=3 [accessed January 10, 2018].
- MUIZNIEKS, N. 2016A, "Diese Pläne sind schlicht illegal", *Tagesschau*. Retrieved from <http://www.tagesschau.de/ausland/eu-tuerkei-fluechtlingskrise-101.html> [accessed May 27, 2018].
- MUIZNIEKS, N. 2016B, "Safeguards needed for EU-Turkey migration deal", *Euractiv*. Retrieved from <https://www.euractiv.com/section/justice-home-affairs/opinion/safeguards-needed-for-eu-turkey-migration-deal/> [accessed July 28, 2018].

- MUIZNIEKS, N. 2016C, “Stop Your Backsliding, Europe”, *The New York Times*. Retrieved from <https://www.nytimes.com/2016/03/15/opinion/stop-your-backsliding-europe.html> [accessed October 9, 2018].
- NEW YORK DECLARATION n.d., “Refugees and Migrants”, UN website. Retrieved from <https://refugeesmigrants.un.org/declaration> [July 3, 2018].
- NUSSBAUM, M. 2019, *The Cosmopolitan Tradition: A Noble but Flawed Ideal*, Harvard: Harvard University Press.
- OWEN, D. 2016A, “In loco civitatis: on the normative basis of the institution of refugeehood and responsibilities for refugees”, F. SARAH; L. YPI (ed.), *Migration in Political Theory*, Oxford: Oxford University Press, 269-89.
- OWEN, D. 2016B, “Refugees, fairness and taking up the slack: On justice and the international refugee regime.” *Moral Philosophy and Politics* 3(2), 141-164.
- OWEN, D. 2019, “Refugees, EU citizenship and the Common European Asylum System. A normative dilemma for EU integration”, *Ethical Theory and Moral Practice* 22: 347-69. <https://doi.org/10.1007/s10677-019-09973-x>
- PAKISTAN n.d., “European Civil Protection and Humanitarian Aid Operations” of the European Commission website.” Retrieved from http://ec.europa.eu/echo/where/asia-and-pacific/pakistan_en [accessed May 7, 2018].
- PROACTIVA OPEN ARMS n.d., “Our Missions.” Retrieved from <https://www.proactivaopenarms.org/en/our-missions> [accessed June 12, 2018].
- SANDELIND, C. 2017, “Costs of refugee admission and the ethics of extra-territorial protection”, *European Journal of Political Theory*. <https://doi.org/10.1177/1474885117738118>
- SINGER, P. 2015, “Escaping the refugee crisis”. Retrieved from <https://www.project-syndicate.org/commentary/escaping-europe-refugee-crisis-by-peter-singer-2015-09?barrier=accesspaylog> [accessed April 19, 2020].
- SINGER, P.; SINGER R. 2010, “The ethics of refugee policy”, J. S. FISHKIN, *Population and Political Theory*. Chichester: Wiley-Blackwell, 285-304.
- SANT’ EGIDIO, 2019, *Dossier sui corridoi umanitari in Italia, Francia, Belgio e Andorra*. Retrieved from <https://www.santegidio.org/downloads/Dossier-Corrido-Umanitari-20190627-web.pdf>.
- TAMPERE EUROPEAN COUNCIL 1999, “Presidency Conclusions”, 15 and 16 October, European Parliament website. Retrieved from http://www.europarl.europa.eu/summits/tam_en.htm [accessed January 19, 2019].
- TURKEY n.d., “European Civil Protection and Humanitarian Aid Operations”, European Commission website. Retrieved from http://ec.europa.eu/echo/where/europe/turkey_en [accessed May 6, 2018].
- UGANDA n.d., “European Civil Protection and Humanitarian Aid Operations”, European Commission website. Retrieved from http://ec.europa.eu/echo/where/africa/uganda_en [accessed May 6, 2018].

- UNHCR 2015, *UNHCR Statistical Yearbook 2014*. Geneva: UNHCR. Retrieved from <https://www.unhcr.org/56655f4d8.html> [accessed April 7, 2018].
- UNHCR 2017A, *Global Appeal*. Geneva: UNHCR. Retrieved from <http://www.unhcr.org/protection/resettlement/593a88f27/unhcr-projected-global-resettlement-needs-2018.html> [accessed November 11, 2018].
- UNHCR 2017B, *Global Trends. Forced Displacement in 2016*. Geneva: UNHCR. Retrieved from <https://www.unhcr.org/5943e8a34.pdf> [accessed May 8, 2018].
- UNHCR 2017C, *UNHCR Statistical Yearbook 2016*. Geneva: UNHCR. Retrieved from <http://www.unhcr.org/statistics/country/5a8ee0387/unhcr-statistical-yearbook-2016-16th-edition.html> [accessed April 7, 2018].
- UNITED NATIONS 2016, “Resolution 71/1. New York Declaration for Refugees and Migrants”, New York. Retrieved from http://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/71/1 [accessed March 24, 2018].
- WALZER, M. 1983, *Spheres of Justice*, New York: Basic Books.
- WELLMAN, C. H.; COLE, P. 2011, *Debating the Ethics of Immigration. Is there a Right to Exclude?*, Oxford: Oxford University Press.
- WELLMAN, C. H. 2016, “Freedom of movement and the rights to enter and exit”, FINE, S; YPI, L. (ed.), *Migration in Political Theory*, Oxford: Oxford University Press, 80-101.

Brúixola filosòfica

VÍCTOR J. LUQUE
Universitat de València

El debat sobre l'estructura causal de la teoria evolutiva: lleis, forces i analogies

*The contemporary debate over the causal structure of evolutionary theory:
laws, forces, and analogies*

Rebut: 15/3/2021. Acceptat: 23/4/2021

Resum: Aquest article analitza l'estructura de la teoria evolutiva i les discussions produïdes als seu voltant en les últimes dècades. En primer lloc, presente l'anomenada "interpretació de forces" o "analogia newtoniana" i la seua estructura basada en l'elaboració de tres lleis fonamentals. Més endavant, mostre les crítiques a aquesta interpretació per diferents autors, en especial l'anomenada "versió estadística" de la teoria evolutiva. Posteriorment, analitze la idoneïtat de la interpretació de forces i elabore una reformulació de la mateixa que permet explicar el seu sorgiment i la seua utilitat dins de la literatura biològica.

Abstract: This article analyses the structure of evolutionary theory and the debates emerged in the last decades regarding its causal status. First, I present the so-called "force interpretation" or "Newtonian analogy," and its structure based on three fundamental laws. Second, I outline several criticisms on this interpretation, specially by the so-called "statistical view" of evolutionary theory. Finally, I analyse the suitability of the force interpretation and elaborate a new formulation of this interpretation. This allows me to explain the emergence of this interpretation in the past and its current utility in the biological literature.

Paraules clau: teoria evolutiva, estructura causal, interpretació de forces, visió estadística, forces evolutives, equació de Price.

Keywords: evolutionary theory, causal structure, force interpretation, statistical view, evolutionary forces, Price equation.

LA BIOLOGIA EVOLUTIVA és una disciplina científica que estudia l'origen i diversitat biològica i el seu canvi a través del temps. Per explicar aquests fenòmens, la teoria evolutiva apel·la a diferents processos o causes evolutives. Així, és força habitual en els llibres de text i en els escrits de la majoria dels biòlegs evolutius parlar de *forces evolutives* que actuen sobre una població. Així Gillespie ens diu: “Els genetistes de poblacions ocupen la major part del seu temps en dues coses: descrivint l'estructura genètica de les poblacions o teoritzant sobre les forces evolutives que actuen sobre les poblacions”. I Hartl i Clark reiteren: “existeixen moltes forces en genètica de poblacions que actuen oposant-se entre elles, i és aquesta tensió la que produeix comportaments interessants a nivell poblacional”¹. De la mateixa manera, en diversos manuals i llibres de text podem trobar capítols dedicats a la interacció de les diverses forces evolutives (per exemple, TEMPLETON 2006, cap. 12). L'analogia amb la mecànica newtoniana ha estat fructífera tant en l'elaboració de models matemàtics com en l'estructuració de la teoria evolutiva. Aquesta analogia, encara que de manera embrionària es troba ja en un treball anterior de Michael Ruse (1973), va ser proposada per Elliott Sober (1984, veure Taula 1) en els següents termes:

Totes les possibles causes evolutives poden ser caracteritzades en termes de “efectes esbiaixadors”. La selecció pot transformar les freqüències gèniques, però també ho poden fer la mutació i la migració. (...) Tot açò situa a la teoria evolutiva en un territori familiar: n'és una teoria de forces.²

Sober planteja que la teoria evolutiva és una teoria de forces perquè, de la mateixa manera que les diferents forces de la mecànica newtoniana provoquen canvis en el moviment dels cossos, les forces evolutives provoquen canvis en les freqüències gèniques i/o genotípiques d'una població. Així, la selecció, la deriva genètica, la mutació i la migració serien les forces o causes principals de l'evolució.³ Un exemple simple d'aquesta conceptualització pot observar-se en

¹ “Population geneticists spend most of their time doing one of two things: describing the genetic structure of populations or theorizing on the evolutionary forces acting on populations” (GILLESPIE 2004, 1); “there are many forces in population genetics that act in opposition to one another, and it is this tension that makes for interesting behavior at the population level” (HARTL i CLARK 1997, 294).

² “All possible causes of evolution may be characterized in terms of their “biasing effects”. Selection may transform gene frequencies, but so may mutation and migration. [...] All this is merely to locate evolutionary theory in a familiar territory: it is a theory of forces” (SOBER 1984, 31).

³ Aquestes varien en nombre, introduint-se a vegades altres factors com la recombinació, estructura poblacional, etc., encara que les quatre esmentades són les considerades més canòniques.

la Figura 1, la qual mostra les pressions selectives per al comportament altruista amb guanys additius i alta relació.

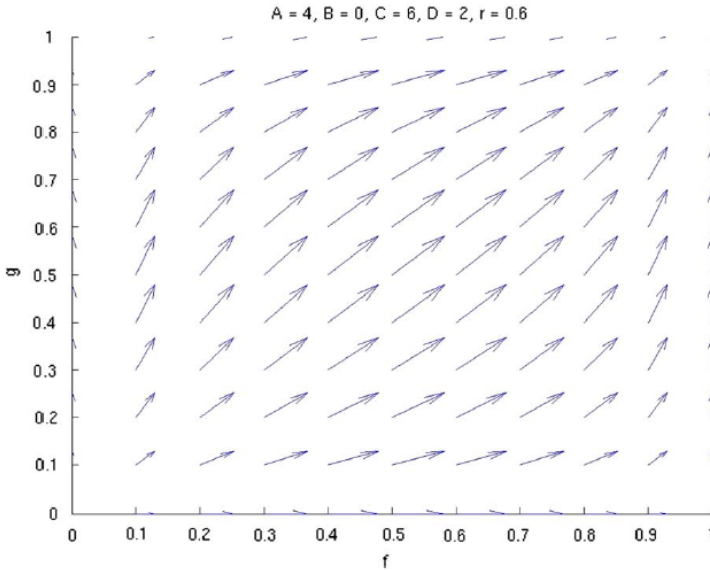


FIGURA 1. Exemple de l'ús de la representació vectorial, a la *Newton*, en biologia evolutiva. L'eix g representa el comportament altruista i l'eix f representa la freqüència poblacional dels altruistes. A, B, C i D representen els paràmetres d'una matriu de pagaments en un joc de Falcó-Colom; i r el coeficient de parentiu (el grau mitjà de relació entre els membres de la població). Seguint la regla de Hamilton, si la raó del cost respecte al benefici és prou alta, relativa al nivell de parentiu o relació de la població, llavors el comportament altruista es fixarà tal com mostren la direcció de les fletxes. Adaptada de Marshall (2009).

Estaríem davant del que Tim Maudlin (2004) anomena 'teories quasi-newtonianes'. Aquestes es caracteritzen per intentar replicar l'estructura de la mecànica newtoniana. D'aquesta manera, la teoria hauria de posseir les següents tres lleis característiques (SOBER 1984):

- Una Llei d'Estat o Força Zero (*Zero-Force Law*). Aquesta ens diria com es comportaria el sistema si no estiguera actuant cap força sobre ell.
- Disposar de Lleis Font (*Source Laws*). Aquestes descriuen les circumstàncies en les quals es produeixen forces.
- Comptar amb Lleis Conseqüencials (*Consequence Laws*). Aquestes descriuen com les forces, ja sorgides en el sistema, produeixen els canvis dins d'aquest.

El principal propòsit de construir teories "quasi-newtonianes" és poder identificar les causes que afecten el sistema d'interès. D'ací la necessitat de posseir una llei de força zero. En la mecànica newtoniana la Primera llei del

moviment o Llei de la inèrcia realitza aquesta funció en establir que tot cos roman en el seu estat de repòs o de moviment uniforme en línia recta, excepte si és compel·lit a canviar el seu estat per forces aplicades (NEWTON 1846 [1687]). De la mateixa manera, la teoria evolutiva tradicionalment ha postulat la seua pròpia llei de força zero amb l'anomenada Llei Hardy-Weinberg. Aquesta llei estableix que si en una població mendeliana no actua cap força evolutiva, aleshores no es produirà cap canvi en ella.⁴

D'aquesta manera, tant la Llei de la inèrcia com la Llei Hardy-Weinberg asseguren un substrat neutral a partir del qual introduir forces, dient-nos com es comportaria el sistema si res ho pertorbara. Un punt important és que la desviació d'aquestes lleis de força zero és una condició suficient per a saber que almenys alguna força està actuant, però no a l'inrevés. Que un sistema complisca l'equilibri que estableix la llei de força zero no implica que cap força estiga actuant. Així, un cos pot estar en repòs perquè dues forces estiguen actuant amb igual magnitud i direcció, però amb sentit oposat sobre ell. De la mateixa manera, una població pot estar en equilibri Hardy-Weinberg però no a causa d'una simple conseqüència de l'aparellament a l'atzar, sinó a la interacció de diferents forces evolutives que es contraposen una a l'altra. Aquestes seria el cas, per exemple, en el model d'equilibri mutació-selecció, on l'acció de la selecció natural en contra d'un al·lel deleteri (per exemple, un que augmenta la mortalitat dels organismes que el tenen) és compensada per una alta taxa de mutació de canvi des de l'al·lel normal cap al deleteri (és a dir, la mutació reintrodueix l'al·lel que la selecció elimina de la població), quedant la població en equilibri i, per tant, sense canvi en les freqüències al·lèliques (la població no evoluciona; veure Gillespie (2004) i Templeton (2006) per a més detalls).

Les Lleis Font (*Source Laws*) proporcionen les circumstàncies de producció de forces. Així, la llei de la gravetat de Newton postula que dos cossos s'atrauen amb una força directament proporcional al producte de les seues masses i inversament proporcional al quadrat de la distància que els separa. En la teoria evolutiva les formulacions de les diferents forces evolutives compleixen una funció anàloga. D'aquesta manera, per exemple, la migració ocorre quan individus són introduïts en una població (immigració) o quan ixen d'aquesta població (emigració).

⁴ Tècnicament, la llei Hardy-Weinberg estableix que en una població diploide i idealment infinita, on hi ha aparellament a l'atzar (població panmíctica) i els individus de la qual són viables i fecunds, la població es mantindrà o tornarà a l'equilibri (és a dir, les freqüències al·lèliques i genotípiques es mantindran estables) si no actua cap força sobre ella, ja que l'herència mendeliana per si sola no pot canviar les freqüències al·lèliques d'aquella. La demostració en la seua forma més simple es realitza així: amb dos al·lells, A i B, en un mateix locus i sent les freqüències dels quals p i q respectivament, les freqüències dels tres genotips AA, AB i BB seran p^2 , $2pq$ i q^2 respectivament (TEMPLETON 2006).

Les Lleis Conseqüencials (*Consequence Laws*), per la seua part, ens informen de com canviarà un sistema una vegada intervé una força. La segona llei del moviment de Newton juga aquest paper en la mecànica newtoniana, informant-nos que el canvi produït per la força serà igual al producte de la massa de l'objecte per la seua acceleració. Aquesta funció la realitzada tradicionalment en la teoria evolutiva la genètica de poblacions, computant i predient els canvis en les freqüències al·lèliques i genotípiques de les poblacions per l'acció d'alguna de les forces evolutives esmentades anteriorment.

TAULA I

TEORIA EVOLUTIVA COM UNA TEORIA DE FORCES	
MECÀNICA NEWTONIANA	TEORIA EVOLUTIVA
Forces com a causes: les forces de la mecànica causen els canvis de posició (o manteniment) dels cossos.	Forces com a causes: la selecció natural, la deriva, la mutació i la migració són forces (causes) que poden donar com a resultat el canvi en les freqüències gèniques (evolució).
Llei de força zero: Llei de la inèrcia	Llei de força zero: Llei Hardy-Weinberg
Models de combinació i resolució de forces: combinació vectorial.	Models de combinació i resolució de forces: La genètica de poblacions proveeix de models que representen com aquestes forces actuen i com combinar-les.

2. HISTÒRIA DE L'ANALOGIA DE FORCES

Un element fonamental és dilucidar com i per què es va desenvolupar la imatge del que posteriorment s'ha denominat l'analogia newtoniana. La interpretació de forces es va anar construint alhora que s'alçaven les bastimentades de la teoria evolutiva moderna. Un primer pilar és establir l'estat per defecte del sistema. Es dona per fet en la historiografia rebuda que va ser la postulació de la llei Hardy-Weinberg en 1908 per G. H. Hardy i Wilhelm Weinberg de manera independent —el primer ho va fer mitjançant una deducció matemàtica, el primer mitjançant experimentació—, el primer pilar.

Encara que això és, a grans traços, correcte, voldria assenyalar que l'estat per defecte que assenyala aquesta llei —si no hi ha factors actuant sobre una població, aquesta es manté sense canvis— ja es trobava en el mateix Darwin i que va ser defensada anteriorment a 1908. Darwin postula que l'estat per defecte dels organismes és la permanència o manteniment de la forma ancestral,

on l'explicació de la desviació requereix de la introducció de factors que alteren la forma ancestral (CAPONI 2004). L'arbre de la vida s'ha desenvolupat mitjançant descendència amb modificació, i un factor principal d'aquesta desviació de la descendència respecte de les generacions anteriors és la selecció natural. Aquesta conceptualització va seguir fins i tot entre els Mutacionistes de principis de segle xx. Així, en 1902 Bateson i Saunders exposen, en un informe per a la Royal Society sobre els seus descobriments genètics i les implicacions de l'herència mendeliana, aquesta visió:

Seria de gran interès estudiar l'estadística d'aquest tipus de poblacions en la naturalesa. Si el grau de dominància pot ser determinat experimentalment, o captar l'heterozigot, i podem suposar que totes les formes s'emparellen amb igual llibertat i fertilitat, i no hi ha selecció natural pel que fa a les formes al·lèliques, deuria ser possible predir amb precisió les proporcions dels diversos components de la població. Pel contrari, desviaments del resultat calculat ens donarien no poca llum sobre la influència de factors alteradors, selecció i similars.⁵

Bateson i Saunders estan enunciant el que posteriorment es coneixerà com a llei d'equilibri Hardy-Weinberg (STOLTZFUS i CABLE 2014). Encara que no van resoldre els càlculs matemàtics que precisa la demostració d'aquesta llei, sí que van reforçar els fonaments conceptuals sobre els estats naturals o per defecte (o en termes de Sober, estats en les condicions estipulades per la llei de força zero) i van anticipar la seua utilitat. És a dir, l'eixida de l'equilibri indicaria que algun factor o força estaria actuant sobre la població, pertorbant aquest equilibri. El paràgraf citat mostra que tant la idea d'un estat per defecte com de factors que alteren aquest estat (*disturbing factors*) es troben en els fonaments de l'estructura de la teoria evolutiva des dels seus inicis.

La següent qüestió important és aclarir d'on provenen tant el concepte de "força evolutiva" com, en general, l'analogia newtoniana. El començament de l'ús del terme "força evolutiva" no està del tot clar. Els arquitectes de la teoria evolutiva moderna —Ronald Fisher, Sewall Wright i J. B. S. Haldane— molt poques vegades van utilitzar aquest terme, al·ludint la majoria de les vegades a "factors", "agents", o "causes", reforçant al seu torn la idea que la teoria evolutiva va ser conceptualitzada des dels seus començaments com una teoria causal.

⁵ "It will be of great interest to study the statistics of such a population in nature. If the degree of dominance can be experimentally determined, or the heterozygote recognised, and we can suppose that all forms mate together with equal freedom and fertility, and that there is no natural selection in respect of the allelomorphs, it should be possible to predict the proportions of the several components of the population with some accuracy. Conversely, departures from the calculated result would then throw no little light on the influence of disturbing factors, selection, and the like" (BATESON i SAUNDERS 1902, 131).

Fisher només utilitza dues vegades el terme “evolutionary force” en el seu llibre fonamental *The genetical theory of natural selection* (1930, 208 i 243). Per la seua part, Haldane mai no esmenta aquest terme en la seua principal obra *The causes of evolution* (1932). Wright majorment parla de “factors” (1931, 100).

No obstant això, quaranta anys després, en la dècada de 1970, l’ús del concepte “força evolutiva” s’ha generalitzat en la literatura i continua fins als nostres dies. Així ho mostra l’influent llibre de text de James Crow i Motoo Kimura *An Introduction to Population Genetics Theory* (1970) on s’usen àmpliament tant el terme com l’analogia amb la mecànica clàssica. A més, se’ns diu que podem trobar en una població “una situació estable de resultats degut a l’equilibri de diverses forces oposades —mutació, selecció, migració i fluctuacions aleatòries”—.⁶ Els mateixos autors citen dos treballs de 1957 que utilitzen “evolutionary forces” en el títol —L. C. Dunn, “Evidences of evolutionary forces leading to the spread of lethal genes in wild populations of house mice”, i L. Sandler and E. Novitski “Meiotic drive as an evolutionary force”—. Per tant, el terme estava ja tan estés en la meitat de la dècada de 1950 que podia usar-se com a títol en una publicació sense que poguera haver-hi malentesos. És a dir, el parlar de forces (*force-talk*) estava generalitzat.

La llavor d’aquest ús pot rastrejar-se fins a Fisher i Wright i en com van conceptualitzar el seu propi treball teòric (ací seguisc a PLUTYNSKY 2007). Fisher (1930) va estar influenciat per la manera de treballar dels físics, assenyalant les dinàmiques de les freqüències gèniques i de com diferents factors (forces) podrien canviar aquestes freqüències. Aquests factors alteren la posició i la velocitat dels gens augmentant la seua freqüència o eliminant-los de la població. La influència del treball de Fisher en l’obra de Wright pot veure’s en el desenvolupament dels anomenats “paisatges adaptatius” (*adaptive landscapes*). Fisher elaborà una representació del procés d’adaptació mitjançant una imatge denominada model geomètric d’adaptació de Fisher (*Fisher’s geometrical model of adaptation*). Aquesta representa l’eficàcia màxima (*maximum fitness*), associada amb un punt de l’espai fenotípic, representant el fenotip òptim. Un individu que es troba en un altre punt pateix una mutació que canvia de manera aleatòria el seu fenotip. Açò incrementa la seua eficàcia només si el canvi acosta al fenotip cap al punt òptim. El model de Fisher intenta provar que els canvis xicotets tenen major probabilitat d’acostar-se cap a l’òptim que canvis molt grans. En aquest espai multidimensional, la mutació (amb efectes molt xicotets) i la selecció treballen conjuntament per a aconseguir aquest òptim. El model geomètric pot representar-se com un paisatge d’un sol pic (veure Figura 2).

⁶ “a stable situation that results from the balance of various opposing forces —mutation, selection, migration, and random fluctuations” (CROW i KIMURA 1970, 175)—.

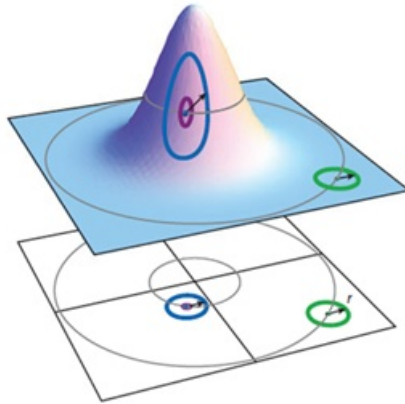


FIGURA 2. Model geomètric d'adaptació de Fisher. L'eficàcia es mostra en l'eix vertical. Per sota tenim una representació en dues dimensions. Adaptada de Tenaillon (2014).

D'aquesta manera, Wright va desenvolupar la idea de la metàfora del paisatge adaptatiu. Els paisatges adaptatius se solen plasmar com un diagrama amb un o dos eixos representant l'eficàcia mitjana d'una població i les freqüències d'uns al·lels concrets. Així, en aquesta imatge trobem turons i valls, on les poblacions recorren la seua superfície, assolint els pics adaptatius o caient a les valls (Figura 3). En aquest marc podem representar diferents forces evolutives movent la població: la selecció l'espenta cap a un pic adaptatiu, la deriva la desplaça de manera aleatòria, etc.⁷

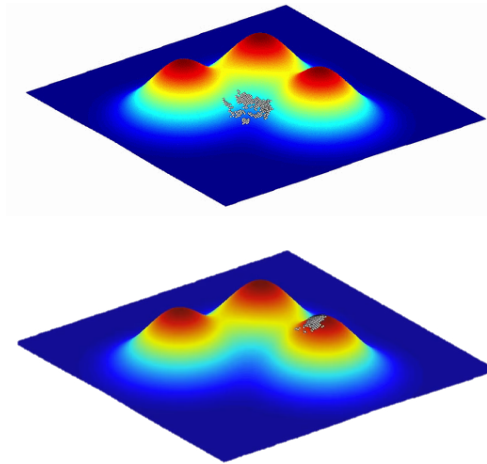


FIGURA 3. Paisatges adaptatius. En primera imatge s'observa una població ascendint cap als pics adaptatius. En un paisatge estàtic, la població només ascendeix fins a un pic adaptatiu, el més pròxim. Simulació realitzada per Randy Olson i Bjørn Østman. Imatge adaptada d'<http://www.randalolson.com/2014/04/17/visualizing-evolution-in-action-dynamic-fitness-landscapes/>

⁷ Per a una defensa del valor heurístic dels paisatges adaptatius, veure Skipper i Dietrich (2012). Per a una visió més crítica, veure Pigliucci i Kaplan (2006, cap. 8).

La interpretació de forces es veu fàcilment arrelada en aquesta imatge tan intuïtiva i fa més fàcil entendre els complexos models matemàtics que Wright i altres teòrics van desenvolupar en aquell període (veure Apèndix A1 per a més detalls). Aquests models (mostrats, visualment, en la Figura 4) estan relacionats amb el Teorema Fonamental de la Selecció Natural desenvolupat per Fisher (1930), on el canvi en l'eficàcia mitjana és proporcional a la variació en eficàcia genètica additiva $\Delta\bar{w} = \frac{V_a(w)}{\bar{w}}$. Com la variància mai no pot ser negativa, la selecció natural sempre incrementarà l'eficàcia mitjana poblacional (si no hi ha variància, l'eficàcia romandrà sense canvis). La interpretació del teorema ha sigut font de disputes (OKASHA 2008).

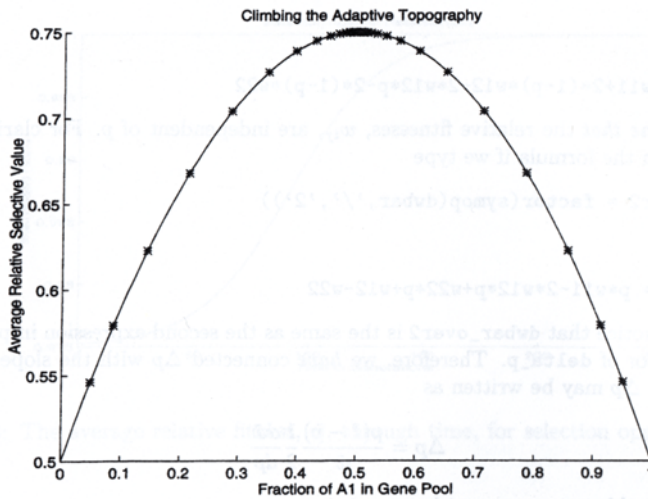


FIGURA 4. Il·lustració de l'acció de la selecció natural, causant que la població ascendisca cap al pic més pròxim de la corba de . Es tracta d'un exemple de superioritat de l'heterozigòtic, el pic del qual es troba en 0.5, predient un polimorfisme estable com a resultat. Adaptada de Roughgarden (1998).

Aquests models bàsics queden englobats sota el paraigua de l'equació de Price o Teorema de Price. Aquesta equació és considerada l'àlgebra de l'evolució (RICE 2004) perquè exposa en un llenguatge matemàtic però simple els diferents processos que interactuen en el fenomen evolutiu. En la seua forma habitual es presenta com a

$$\bar{w}\Delta Z = Cov(w, z) + E(w\Delta z)$$

on $Cov(w, z)$ representa el canvi degut a la supervivència i reproducció, i $E(w\Delta z)$ representa el canvi en la transmissió entre pares i fills. He defensat (LUQUE 2017 i 2018) que l'equació de Price pot jugar un paper fonamental en la teoria evolutiva. Més concretament, l'equació de Price pot ser considerada l'equivalent a la segona llei de Newton per a la teoria evolutiva. En primer

lloc, es una representació abstracta i general del fenomen evolutiu en tant que no especifica quines forces actuen sobre la població, quantes són, ni quines formes adopten. L'equació de Price únicament ens diu com computar-les una vegada apareixen en el sistema. Així, aquesta equació seria la llei conseqüencial de la teoria evolutiva, adoptant diferent formes depenent de les forces evolutives que estiguen actuant en el sistema. Així, si hi ha selecció direccional en la població (es a dir, la selecció afavoreix un fenotip extrem de la població) l'equació de Price adopta la següent forma $\bar{w}\Delta\bar{g} = \beta_{w,g}Var(g)$. Si hi ha mutació, pot adoptar la següent forma $\bar{w}\Delta\bar{g} = E(w\Delta_m g)$. O si diverses forces, com la selecció i la deriva, estan actuant alhora, aleshores l'equació de Price pot adoptar la següent forma $\bar{w}\Delta\bar{z} = Cov(w^*, z') + Cov(\delta, z')$. Així, estaríem davant del que Thomas Kuhn denominava un esquema o esbós de generalització (*generalization-sketch*), proporcionant un marc unificat per als investigadors, que els permet desenvolupar models específics (veure LUQUE 2017 per a més exemples). Aquest paper aplicat a l'equació de Price la connecta amb altres esquemes generalitzadors com la segona llei del moviment de Newton, on ambdues serien lleis conseqüencials dins de les seues respectives teories. Açò dóna un punt a favor de l'analogia de forces, en particular mostrant que la teoria evolutiva compta amb la seua pròpia llei conseqüencial equivalent a la que podem trobar en la mecànica newtoniana, en comptes d'un conglomerat de models simples i a vegades poc interconnectats entre si com els que trobem en genètica de poblacions. Això sens dubte serveix com a recurs i reforç als defensors de la interpretació de forces (més sobre açò en la secció Conclusions).

Retornant al recorregut històric, una part fonamental de la història es troba, possiblement, en la Teoria dels Equilibris Desplaçats (*Shifting Balance Theory*) de Sewall Wright. En aquesta teoria, Wright va estipular tres fases diferents. En una primera fase (fase exploratòria), la deriva genètica actua sobre els diferents demes de la població (subdivisió poblacional) obtenint múltiples combinacions gèniques que exploren diferents pics adaptatius. En una segona fase (fase de selecció masal) la selecció natural expandeix la combinació gènica favorable en els demes que obtenen aquesta combinació, aconseguint un pic adaptatiu, augmentant de grandària i disseminant més emigrants als altres. En una tercera fase (fase de selecció interdèmica), aquestes combinacions gèniques s'expandeixen a la metapoblació mitjançant selecció interdèmica. En el seu model geomètric, Fisher considerava un o pocs pics molt similars entre si, on una població de gran grandària i amb aparellament aleatori està sotmesa principalment a l'acció de la selecció. Wright, en canvi, atorga molta més importància al paper de la deriva genètica i a l'existència de diferents pics adaptatius dins d'un mateix paisatge adaptatiu. Wright afirma (1931):

S'hi suggereix, en conclusió, que les diferents situacions estadístiques esperades entre les espècies naturals son adients per a tindre en compte els diferents tipus de processos evolutius que s'han descrit, i que aquestes condicions en la naturalesa, en particular, són sovint del tipus que produeixen un estat d'equilibri entre tendències, sobre les quals un nombre indefinit de processos evolutius depenen.⁸

Pot ser que no use la paraula força (*force*), però es necessita molt poc, un xicotet salt, per a passar a pensar en l'evolució com una teoria de forces una vegada es té la visió de Wright.

Per a la consolidació de la interpretació de forces, l'obra de Theodosius Dobzhansky *Genetic and the Origin of Species* (1937) va jugar un paper important. En ella Dobzhansky va popularitzar la definició d'evolució com a canvi en les freqüències gèniques, establint les proporcions Hardy-Weinberg com una regla bàsica. Al seu torn, Dobzhansky explicita la visió dinàmica de la teoria evolutiva, on els diferents agents funcionen com a forces que canvien les freqüències gèniques:

L'evolució es un procés de canvi o moviment. La descripció de qualsevol moviment pot ser dividit en dos parts: estàtica, la qual tracta sobre les forces que produeixen un moviment o l'equilibri d'aquestes forces; i la dinàmica, la qual tracta sobre el moviment en si i l'acció de les forces que el produeixen. Seguint aquest esquema, deuríem discutir, primer, els factors que produeixen els canvis en la composició genètica de les poblacions (estàtica evolutiva), i segon, les interaccions d'aquestes forces en la formació i desintegració de races i espècies (dinàmica evolutiva).⁹

En aquesta tercera edició de l'obra de Dobszhansky el llenguatge de forces (*force-talk*) es fa explícit, consolidant la interpretació de forces i explicant per què apareixen articles com els de L. C. Dunn, "Evidences of evolutionary

⁸ "It is suggested, in conclusion, that, the differing statistical situations to be expected among natural species are adequate to account for the different sorts of evolutionary processes which have been described, and that, in particular, conditions in nature are often such as to bring about the state of poise among tendencies on which an indefinitely continuing evolutionary processes depends" (WRIGHT 1931, 158).

⁹ "Evolution is a process of change or movement. Description of any movement may be divided into two parts: statics, which treats of the forces producing a motion and the equilibrium of these forces, and dynamics, which deals with the motion itself and the action of forces producing it. Following this scheme, we shall discuss, first, the factors which bring about changes in the genetic composition of populations (evolutionary statics), and second, the interactions of these forces in race and species formation and disintegration (evolutionary dynamics)" (DOBZSHANSKY 1951, 18).

forces leading to the spread of lethal gens in wild populations of house mice”, i L. Sandler and E. Novitski “Meiotic drive as an evolutionary force”, usant aquest llenguatge en aquesta dècada.

3. CRÍTICA DE LA VISIÓ ESTADÍSTICA

Aquesta visió dominant, no obstant això, s’ha vist desafiada en les últimes dues dècades. Diversos autors (WALSH *et al.* 2002; MATTHEN I ARIEW 2002, 2005, 2009; PIGLIUCCI i KAPLAN 2006; WALSH 2007; WALSH *et al.* 2017) han defensat que l’analogia de la teoria evolutiva amb la mecànica newtoniana no és sostenible. Per això postulen una nova visió denominada estadística (VE) —en contraposició al que s’ha denominat visió tradicional o dinàmica (VD)— on el procés evolutiu i les seues diferents parts (la selecció, la deriva, etc.) són un resultat merament estadístic, i per tant inseparables les unes de les altres. Defensen que no és possible, com en la mecànica newtoniana, separar les (suposades) diferents forces que actuen sobre les poblacions. Per a aquests autors, si hi ha una analogia correcta amb alguna branca de la física, no és amb la mecànica newtoniana, sinó amb la termodinàmica estadística. Fenòmens com l’increment de l’entropia no són explicats mitjançant processos causals sinó com meres tendències estadístiques. D’aquesta manera, les anomenades forces evolutives haurien d’entendre’s com una tendència estadística a nivell poblacional, abandonant qualsevol pretensió de rol causal d’aquestes. És a dir, serien simples agregats matemàtics d’esdeveniments a nivell individual (naixements, morts, reproduccions, etc.) (veure Taula 2).

TAULA 2

VISIÓ DINÀMICA	VISIÓ ESTADÍSTICA
Forces com a causes.	L’evolució és un conjunt de resultats constatables a nivell poblacional.
Estat de força zero: Llei Hardy-Weinberg.	No existeix l’estat de força zero. L’estat per defecte és la deriva.
Models de combinació i resolució de forces evolutives.	No es poden combinar perquè aquestes forces evolutives no existeixen.

En primer lloc ataquen la mateixa existència d'una Llei de força zero en la teoria evolutiva. Sostenen que la teoria evolutiva, a diferència de la mecànica newtoniana, depèn d'un substrat físic (la genètica mendeliana) per a funcionar. Aquesta dependència fa difícil imaginar, per a aquests autors, com podran separar-se les diferents forces i diferenciar-les de les produïdes pel substrat. Al seu torn, la llei Hardy-Weinberg no és aplicable a organismes de reproducció asexual, per la qual cosa no podria donar compte de la història evolutiva en un interval de dos mil milions d'anys. En tot cas, podem apel·lar al fet que la VD, favorable a l'analogia de forces, defensa la idea general que no existeix canvi en les freqüències gèniques i/o genotípiques si no actua alguna força sobre la població (més sobre açò en la secció 4). Per contra, per a la VE les poblacions, lluny d'estar en equilibri, es troben sotmeses contínuament a errors de mostreig (*sampling error*), coneguts com deriva genètica, les causes de la qual a nivell individual són indistingibles de les produïdes per la selecció natural. Per això resulta difícil per a aquests autors veure com es podria afegir la deriva genètica com una causa quan, per definició, manca de direcció (aquest tema es reprendrà en la subsecció 4.4).

Una de les característiques fonamentals de la mecànica newtoniana és que és possible diferenciar entre les diferents forces que actuen en un sistema i combinar-les. Aquesta resolució de forces permet comparar la força de la gravetat amb la força electroestàtica que, encara que apel·len a diferents unitats (massa i càrrega elèctrica, respectivament), són expressables vectorialment. No obstant això, aquest no seria el cas per a les forces evolutives. Centrant-se en la selecció natural, la VE defensa que la visió tradicional de l'eficàcia biològica (*fitness*)¹⁰ és eminentment comparativa, específica en contextos restringits i per això impossible d'usar els seus resultats concrets en altres contextos, obligant a estimar-la estadísticament pels seus efectes (el nombre actual de descendents). No és possible usar una llei d'addició com en la mecànica newtoniana, no sols respecte a l'eficàcia biològica sinó amb la resta de suposats factors que ens permeten dir quant canvi evolutiu es correspon a la selecció, quant a la deriva, quant a la mutació, etc. (veure Figura 5).

¹⁰ Els autors la denominen eficàcia intrínseca o vernacla (*vernacular fitness*) en contraposició a l'eficàcia predictiva (*predictive fitness*). La primera és l'habilitat d'un organisme per a sobreviure i reproduir-se (una visió causal) amb un caràcter comparatiu però no quantitatiu. L'eficàcia predictiva, en canvi, seria l'usada en genètica de poblacions: seria una mesura estadística del canvi evolutiu que prediu el canvi de freqüències gèniques, la taxa de creixement (positiva o negativa) esperada de les parts d'una població dividides sobre la base de la variació en algun gen o tret. Aquesta eficàcia biològica, per tant, no causa el creixement, sinó que és el creixement.

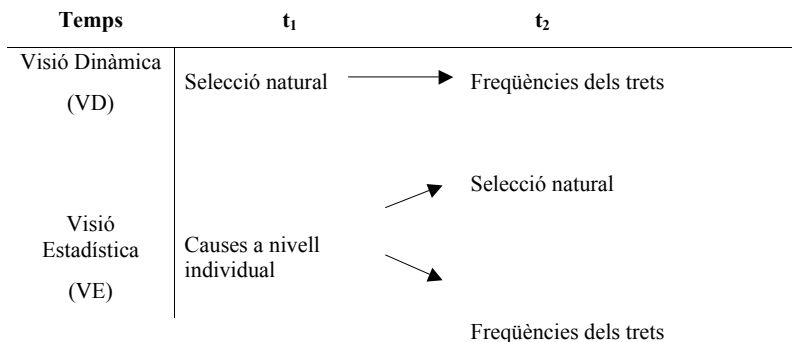


FIGURA 5. Adaptada de Shapiro i Sober (2007).

Per això els defensors de la VE apel·len al fet que els models que aporta la genètica de poblacions són sumaris estadístics de fets a nivell individual, agrupant-los en conjunts (*ensembles*) i fent emergir d'aquesta manera les tendències poblacionals. Aquestes, lluny de ser la causa del canvi, són una conseqüència del mateix i alhora són instanciables en qualsevol substrat. Per aquesta raó consideren que, com va intuir Fisher (1930)¹¹, el paral·lelisme més pròxim amb la física no és la mecànica newtoniana, sinó la mecànica o termodinàmica estadística la qual té una aplicabilitat universal —en tractar de conjunts— sense descriure la constitució de les entitats ni les seues dinàmiques.

Un avantatge que tindria VE enfront de VD és que aclariria diferents debats. Un d'ells seria sobre la visió negativa o positiva de la selecció natural respecte a la possessió dels trets dels individus (DARWIN 1859; DE VRIES 1904; AYALA 1970; SOBER 1984; GOULD 2002) o dit d'una altra manera, si la selecció crea al més eficaç o simplement el preserva. La VE dissoldria aquest problema ja que la selecció no té cap poder causal sobre l'individu, de la mateixa

¹¹ Certament, Ronald Fisher (1930) apel·la dues vegades a l'analogia de la mecànica estadística en la seua famosa obra. La primera és quan compara la teoria cinètica dels gasos amb la genètica mendeliana (FISHER 1930, 11). La segona quan compara el seu Teorema Fonamental de la Selecció Natural amb la segona llei de la termodinàmica (*ibid*, 36). Curiosament, tot seguit (37) Fisher enumera cinc diferències fonamentals entre el seu Teorema i la Segona llei de la termodinàmica: "(1) The systems considered in thermodynamics are permanent; species on the contrary are liable to extinction, although biological improvement must be expected to occur up to the end of their existence. (2) Fitness, although measured by a uniform method, is qualitatively different for every different organism, whereas entropy, like temperature, is taken to have the same meaning for all physical systems. (3) Fitness may be increased or decreased by changes in the environment, without reacting quantitatively upon that environment. (4) Entropy changes are exceptional in the physical world in being irreversible, while irreversible evolutionary changes form no exception among biological phenomena. Finally, (5) entropy changes lead to a progressive disorganization of the physical world, at least from the human standpoint of the utilization of energy, while evolutionary changes are generally recognized as producing progressively higher organization in the organic world."

manera que la teoria dels gasos no pot explicar per què una partícula concreta té la velocitat que té. L'altre debat seria sobre les "unitats de selecció", on pot actuar la selecció i, més important, quina seria la unitat fonamental: gens, individus, grups, etc. (LEWONTIN 1970; WILLIAMS 1992; OKASHA 2006). Sota la VE, la selecció ha quedat reduïda a un recompte o resum (*bookkeeping*), podent aplicar-se a qualsevol nivell (gènic, individual, etc.) sense apel·lar al seu poder causal.

Des que va ser formulada, la VE ha creat una enorme quantitat de literatura sobre aquest tema. En el que queda d'article analitzaré diversos dels problemes plantejats per la VE. No obstant això, els debats i les seues ramificacions han augmentat de manera incessant fent impossible entrar detalladament en tots i cadascun d'ells. Una mostra de les diferents posicions dels autors en contesa les he exposades en la Taula 3, resumint els elements principals que es debaten: nivell causal, poder causal de l'eficàcia biològica, entendre la selecció i la deriva com a parts del mateix procés o com a diferents processos, i marcs causals. El meu principal objectiu és mostrar les línies generals d'aquests debats, permetent a lector navegar entre ells i, en última instància, abordar-los proposant possibles solucions.

TEORIA DE L'EVOLUCIÓ

	ESTADÍSTICS	CAUSALISTES		NIVELL CAUSAL		EFICÀCIA BIOLÒGICA		SELECCIÓ/ DERIVA		MARC CAUSAL			
		Interpretació de forces	Altres	Individual	Poblacional	Causal	No causal	Un procés	Diferents processos	Intervencionista	Contrafactual	Probabilístic	Difference-maker
Abrams			*		*	*		*		*		*	
Arrew	*			*			*	*		*			
Bouchard			*	*		*		*	*				
Brandon		*			*	*		*		*			
Clatterbuck		*			*	*		*		*			
Filler		*			*	*		*		*			
Gildenhuis		*		*		*		*		*			
Glennan		*		*		*		*		*			
Glymour		*		*		*		*		*		*	
Huneman		*			*			*		*		*	*
Kaplan	*			*		*		*		*		*	
Lewens	*		*	*		*		*		*			
Matthen	*			*		*		*		*			
McShea		*			*	*		*		*			
Millstein		*			*	*		*		*		*	
Otsuka		*		*		*		*		*		*	
Pence		*			*	*		*		*		*	
Pigliucci	*			*		*		*		*		*	
Ramsey		*		*		*		*		*			
Rosenberg		*		*		*		*		*			
Sarkar		*		*		*		*		*		*	*
Sober		*		*		*		*		*		*	
Stephens		*		*		*		*		*		*	*
Walsh	*			*		*		*		*		*	

TAULA 3. Panorama de las diferents postures que han adoptat alguns autors destacats respecte als temes derivats de la discussió entre estadístics i causalistes. Els autors que defensen totes dues opcions en una dicotomia és perquè han canviat la seua visió al llarg dels anys o perquè ambdues són compatibles. Si la secció apareix en blanc és perquè no s'han pronunciat respecte a eixe tema.

4. DEFENSA DE L'ANALOGIA DE FORCES

Com hem vist, hi ha un desafiament a la VD de la teoria d'evolutiva l'eix principal de la qual és atacar l'analogia amb la mecànica newtoniana. En els primers apartats defensaré que l'analogia és perfectament aplicable a la teoria evolutiva pel que fa a la selecció natural, a la mutació i a la migració. A la deriva genètica, per ser el punt més controvertit, li reservem els apartats 4.4 i 4.5.

4.1 Descomposició i composició de forces

Un dels avantatges de la combinació vectorial newtoniana és que permet descompondre el vector final en els diferents vectors que el componen. La pregunta que se'ns imposa és si podem separar la selecció natural, la mutació i la migració en els seus diferents components i si després podem combinar-los de nou. Comencem per la selecció. No és difícil trobar en els llibres de text el procés de selecció descompost en diferents fases del cicle vital (FONTDEVILA i MOYA 1999; FUTUYMA 2013). Així, els components de la selecció en organismes de reproducció sexual els podem dividir entre la supervivència i la reproducció. Això ja ens permetria contraargumentar a la VE, però podem seguir amb la divisió de la selecció en: selecció en viabilitat, selecció sexual, selecció en fecunditat i selecció gamètica (veure Figura 6).¹²

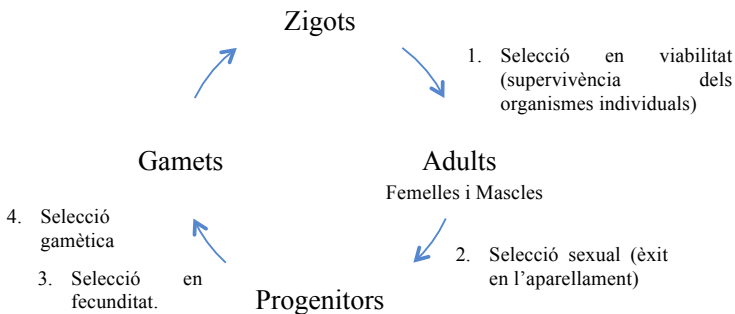


FIGURA 6. Components de la selecció en organismes de reproducció sexual. Adaptada de Futuyma (2013).

¹² Ací cal considerar dos elements. Primer, la divisió dels components de la selecció que afecten l'eficàcia no són universals —per exemple, els bacteris manquen d'èxit en l'aparellament (selecció sexual)—. Segon, es corre el risc de l'arbitrarietat en la subdivisió (ORR 2009). No obstant això, aquest risc d'arbitrarietat en la subdivisió també pot donar-se en mecànica clàssica.

Així mateix, la migració pot descompondre's entre la immigració i l'emigració, entre el nombre d'individus que exporta una població a una altra i els individus que rep al seu torn d'una altra població. De la mateixa manera, la mutació pot ser dividida: si tenim dos al·lells alternatius en un locus, A i B , podem obtenir la taxa de mutació de l'al·lel A cap al B i la taxa de mutació de l'al·lel B per a convertir-se en l' A (STEPHENS 2010). No sembla haver-hi problema a descompondre i compondre de nou aquestes forces evolutives. A més podem veure com interaccionen entre elles. Un model clàssic de genètica de poblacions que mostra aquestes interaccions és el d'equilibri mutació-selecció explicat, breument, en la secció 1 (veure Apèndix A2 per a més detalls). Aquest model simple, que contesta parcialment a la crítica de VE, pot complementar-se amb un altre que incloga també a la migració. Imaginem l'evolució de dos al·lells, A i B , en una població haploide i generacions discretes, on la freqüència de l'al·lel A en la següent generació (generació 2) és el producte del seu fitness, w_A , per la freqüència en la primera generació, p_1 . Així la seua freqüència en la generació 2 serà $p_2 = w_A p_1$. Aplicant el mateix raonament per a l'al·lel B la freqüència del qual serà q_1 , obtenim que la seua freqüència en la generació 2 serà $q_2 = w_B q_1$. La taxa de mutació des de l'al·lel A cap al B serà μ i des de l'al·lel B cap a l'al·lel A serà ν . Per part seua, la taxa de pèrdua de l'al·lel A per l'emigració serà m_{1A} i el guany del mateix per la immigració serà m_{2A} i les taxes de pèrdua i guany per l'emigració i immigració de l'al·lel B seran m_{1B} i m_{2B} respectivament. D'aquesta manera podem combinar el canvi genètic produït per l'acció de la selecció, la mutació i la migració per a la generació 2, per a tots dos al·lells, mitjançant

$$p_2 = w_A p_1 + p_1(1 - \mu) + (1 - p_1)\nu - m_{1A} + m_{2A}$$

$$q_2 = w_B q_1 + q_1(1 - \nu) + (1 - q_1)\mu - m_{1B} + m_{2B}$$

El model¹³ no és completament additiu, com semblen defensar Brandon i Ramsey (2007), en tant que el que s'ha calculat són les freqüències absolutes, per la qual cosa caldria normalitzar dividint per $p_2 + q_2$ per a obtenir les freqüències relatives (HITCHCOCK i VELASCO 2014). La falta d'una addició completa podria ser un punt a favor de la crítica de la VE segons la qual les forces evolutives no són tals perquè no permeten aquesta addició i, per tant, no serien capaços de separar-se i observar el seu poder causal amb independència les unes de les altres. Hitchcock i Velasco (2014) han assenyalat encertadament que aquests requisits tan estrictes no són assolits tampoc en diverses parts de la mecànica newtoniana. Així, en les forces electroestàtiques els cossos tenen una càrrega elèctrica però també una massa, per la qual cosa creen una força gra-

¹³ És un model presentat per Brandon i Ramsey (2007). Hi ha un model equivalent que substitueix la migració pel sistema d'aparellament en Templeton (2006, 377).

vitacional obtenint dues forces que sorgeixen al mateix temps; i de la mateixa manera ocorre amb la força de fricció, que sempre va acompanyada d'una força normal. L'addició vectorial, en sentit estricte, no pot aplicar-se a tots els sistemes newtonians. Per exemple, entre molècules polaritzades movent-se en un camp elèctric, on el moment elèctric induït de la molècula depèn de la localització de la molècula en aquest camp, i aquesta depèn al seu torn de la localització de les altres molècules i els seus camps (CORBEN i STEHLE 1994, 29).

4.2 Magnitud, direcció i sentit

En la mecànica newtoniana, els trets que caracteritzen als vectors són magnitud, direcció i sentit. La qüestió és si les forces evolutives posseeixen aquestes tres característiques. Hem vist en l'apartat anterior com la mutació tenia una magnitud —mesurada com la taxa de mutació respecte d'un al·lel feia l'altre al·lel— una direcció i un sentit —el canvi des d'un al·lel feia l'altre—. La migració també compta amb una magnitud —mesurada com la taxa de pèrdua o guany d'un al·lel per la seua emigració o immigració— i, a més d'una direcció i sentit —emigrant cap a una altra població o immigrant a la pròpia—. En el cas de la selecció natural comptem amb una magnitud clara, el coeficient de selecció, però respecte a la seua direcció i sentit hem de ser una mica més específics.

En genètica quantitativa la selecció sol dividir-se en tres possibles formes: direccional, estabilitzadora i disruptiva (GILLESPIE 2004; TEMPLETON 2006). La selecció direccional ofereix amb claredat el que necessitem ja en el seu mateix nom, on el fenotip extrem de la població es veu afavorit, dotant d'una direcció clara (vectorialment: sentit) al canvi poblacional. La selecció estabilitzadora afavoreix al fenotip intermedi en detriment dels fenotips extrems, mentre que la selecció disruptiva afavoreix als fenotips extrems enfront dels intermedis. En aquests dos casos, la direccionalitat de la selecció no resulta tan clara i podria posar en perill la seua representació vectorial, excepte si pensem en els fenotips de la població com subpoblacions que tenen el seu propi sentit. Això no és diferent al que ocorre amb la gravetat. Imaginem que en el nostre sistema solar es troben dos objectes amb igual massa i a la mateixa distància del Sol, però en costats oposats d'aquest. Si abstraïem els efectes de qualssevol altres objectes sobre aquest sistema, llavors les dues masses experimentaran una força amb la mateixa magnitud i direcció però amb sentit oposat. De la mateixa manera, en una població sotmesa a selecció estabilitzadora dues subpoblacions poden trobar-se equidistants del punt seleccionat però en costats oposats (BRANDON 2006).

La idea bàsica és que un procés com el de la selecció afavoreix a uns tipus concrets en la població en detriment d'uns altres —és a dir, actua de manera

discriminada— mentre que processos com la deriva, per definició, no afavoreixen a cap en concret —és a dir, actua de manera indiscriminada (MILLSTEIN 2002)—. D’aquesta manera, les diferents formes de selecció poden representar-se en els paisatges adaptatius, indicant quin tipus de fenotip està sent afavorit per la selecció (veure Figura 7).

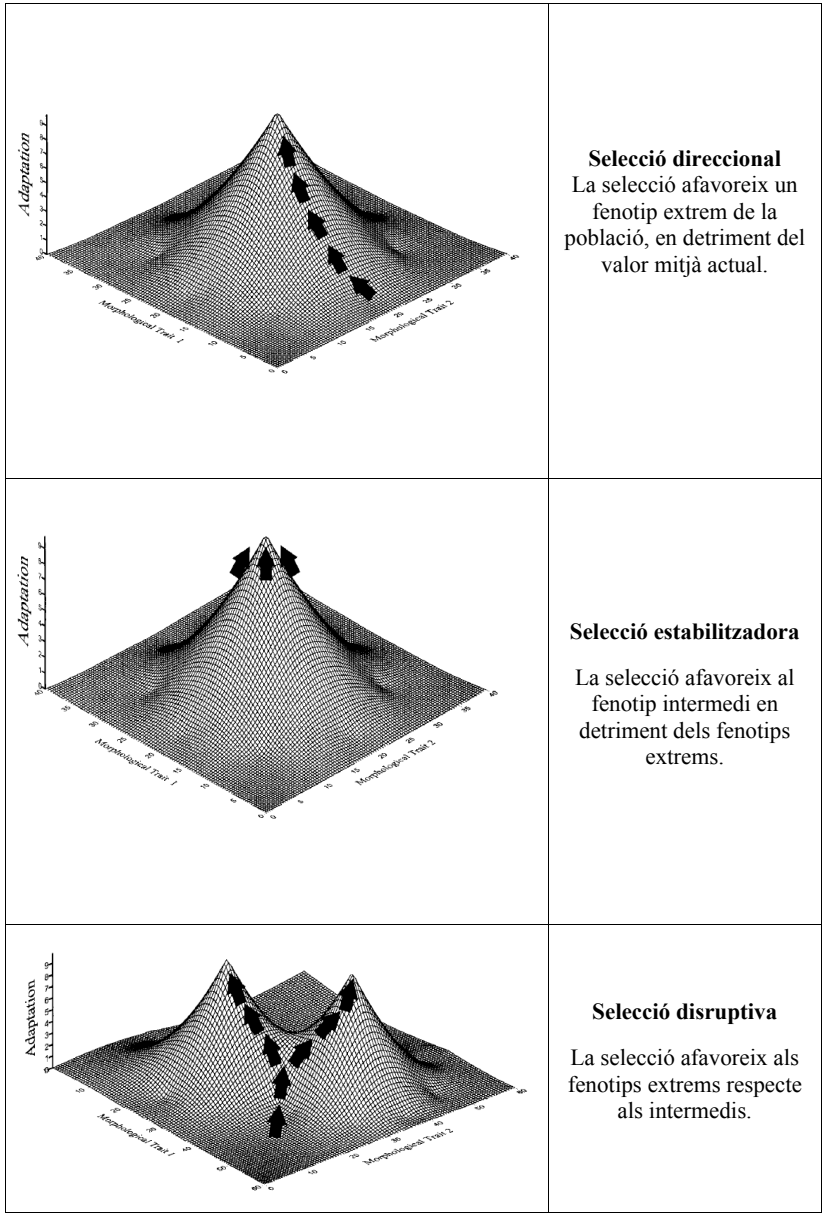


FIGURA 7. Adaptada de McGhee (2007).
Representació de tres formes de selecció en genètica quantitativa.

Les forces evolutives satisfan les propietats exigibles a un vector (magnitud, direcció i sentit) i a més permeten computar-les sense major problema. El seu comportament és anàleg a les forces newtonianes i, per això, l'analogia és il·luminadora. Així ho conceptualitza, per exemple, pel biòleg matemàtic Sean H. Rice: “Podem representar els seus efectes en termes de camps vectorials sobre un espai de freqüències al·lèliques [...] Combinar aquests processos resulta fàcil, en tant que, per a qualsevol valor de p podem simplement afegir els dos camps vectorials per aconseguir els efectes conjunts d'ambdós processos. En aquest sentit, selecció i mutació es comporten com les forces en física”¹⁴ (veure Figura 8). El model d'equilibri mutació-selecció és anàleg al comportament d'una bola de billar que es troba en repòs perquè dues forces amb igual magnitud però amb sentit oposat estan actuant sobre ella. De la mateixa manera, la mutació s'oposa a la selecció al no permetre que s'elimine l'al·lel recessiu, reintroduint-lo cada vegada.

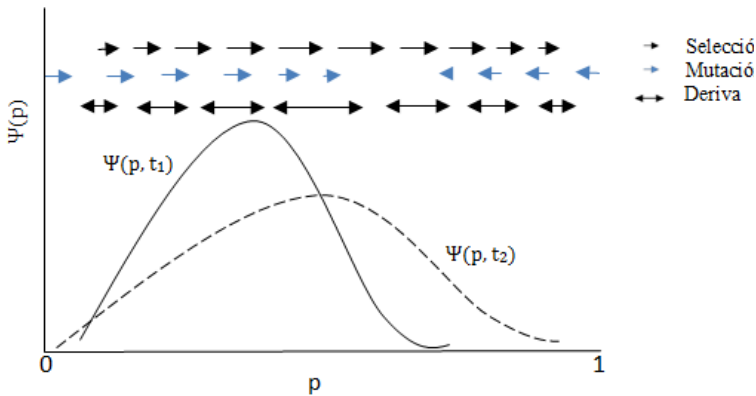


Figura 8. Adaptada de Rice (2004). En ella es mostra el desplaçament en la distribució de les freqüències al·lèliques de la població pels efectes direccionals de la selecció i la mutació, i el seu aplanament pels efectes de la deriva. Veiem que la deriva manca de direcció, ja que el seu canvi esperat en cada punt és zero.

4.3 Causalitat: processos i productes

Abans de seguir amb l'analogia de forces, és necessari una anotació respecte al caràcter causal d'aquestes. Diversos autors han defensat (STEPHENS 2004; MILLSTEIN 2006; SHAPIRO i SOBER 2007) que la VE perd l'element causal de les forces evolutives o processos evolutius perquè han centrat la seua atenció en el

¹⁴ “We can represent their effects in terms of vector fields over the space of allele frequencies [...] Combining these processes becomes easy, in that, for any value of p we can simply add together the two vector fields to get the joint effects of the two processes. In this sense, selection and mutation behave like forces in physics” (RICE 2004, 130-1).

producte o resultat d'aquestes forces i no en el procés. Aquesta ambigüitat entre procés i producte la podem trobar en la parla dels mateixos biòlegs i fins i tot en les definicions dels llibres de text. Si es pren al peu de la lletra la definició d'un procés pel seu resultat es corre el risc de caure en la fal·làcia operacionalista (STEPHENS 2004). Futuyma defineix la selecció natural com “la supervivència i/o reproducció diferencial de les classes d'entitats que difereixen en una o més característiques”; i Freeman i Herron l'entenen com “una diferència, en mitjana, entre la supervivència i fecunditat dels individus amb certs fenotips comparats amb individus amb altres fenotips”.¹⁵ Totes dues fórmules defineixen a la selecció natural pel seu efecte —la reproducció diferencial, la qual també és l'efecte de la deriva— i no sembla albirar-se cap contingut causal en elles. No obstant això, una lectura més detallada permet veure que aquesta visió a-causal, lluny de mantindre's, es converteix en una visió causal de la selecció. Per exemple, Futuyma afirma que “la deriva genètica i la selecció natural són les dues causes més importants de substitució al·lèlica —és a dir, de canvi evolutiu— en les poblacions”.¹⁶ A més a més, Futuyma sosté igual que altres autors (LANDE i ARNOLD 1983; FAIRBAIN i REEVE 2001) que cal diferenciar entre la selecció natural i l'evolució per selecció natural. La selecció natural com a tal es dona en condicions intra-generacionals, mentre que l'evolució per selecció natural seria un procés intergeneracional com a resposta a la selecció intrageneracional i dependent del component hereditari. La resposta a la selecció és el que hem denominat producte o resultat de la selecció. Freeman i Herron, per la seua part, defensen que existeixen quatre mecanismes principals causants de l'evolució: la selecció natural, la deriva, la mutació i la migració.

Aquesta confusió entre el procés i el producte la trobem també en les definicions de la deriva genètica, com quan Roughgarden ens diu que “la deriva genètica es el nom per als canvis en les freqüències gèniques causades per aquest error de mostreig”.¹⁷ A més, això es veu afavorit per l'ús de termes com a *efecte Sewall Wright*, l'*efecte Hedgescock*, etc., per a la deriva. Aquesta ambigüitat entre procés i producte, que sembla ser més comú en anglés,¹⁸ ha de ser aclarida

¹⁵ “The differential survival and/or reproduction of classes of entities that differ in one or more characteristics” (FUTUYMA 2013, G-9); “a difference, on average, between the survival or fecundity of individuals with certain phenotypes compared with individuals with other phenotypes” (FREEMAN i HERRON 2007, 803)

¹⁶ “Genetic drift and natural selection are the two most important causes of allele substitution —that is, of evolutionary change— in populations” (FUTUYMA 2013, 258).

¹⁷ “Genetic drift is the name for changes in gene frequency caused by this sampling error” (ROUGHGARDEN 1996, 57-8).

¹⁸ Alguns termes en anglés (“adaptation”, “get married”, etc; veure SOBER 1984, 196) fan referència tant al procés com al producte d'aquest procés. D'aquesta manera casar-se conclou en la condició de casat, o un tret adaptatiu conclou sent (sempre que la selecció natural fixe aquest tret adaptatiu en la població en les següents generacions) una adaptació (el producte, el qual fa referència a la seua història i no a la seua funció actual).

si volem veure el component causal de les forces evolutives. Si la nostra atenció es dirigeix solament cap al resultat o efecte (canvi en les freqüències gèniques) dels processos evolutius, no sols perdem el component causal d'aquests, sinó que se'ns escapen importants característiques evolutives. Tornem al nostre exemple del model d'equilibri mutació-selecció. En ell postulàvem que en una població un al·lel, *A*, era afavorit per la selecció en contra de l'al·lel *B*; mentre que la taxa de mutació des de l'al·lel *A* cap al *B* excedia a la taxa de mutació des de l'al·lel *B* cap a l'*A*. Passades diverses generacions la població arribava a un equilibri, on ja no hi havia canvi en les freqüències al·lel·liques (ja no hi ha evolució). Segons els defensors de la VE, i si ens basem només en les definicions que se centren en el resultat de les forces evolutives, en la població no hi ha cap procés actuant ja que no hi ha evolució. No obstant això, ni la selecció ni la mutació han deixat d'actuar sobre la població, el procés d'ambdues continua donant-se però el seu efecte (l'evolució) no es produeix perquè totes dues forces, igual que en la mecànica newtoniana, es compensen (s'anul·len) (STEPHENS 2004; SHAPIRO i SOBER 2007).

Certament, la crítica de la VE ha obligat a replantejar de quina forma establim el poder causal dels principals processos evolutius. Especialment ha impel·lit a diversos autors a posicionar-se respecte al nivell on es troba el component causal de les dues forces —selecció i deriva— en les que s'ha centrat el debat. Alguns autors (BOUCHARD i ROSENBERG 2004; ROSENBERG i BOUCHARD 2005) han negat per complet la validesa de l'argument de la VE situant tota l'acció causal al nivell dels individus —i per tant tallant d'arrel la concepció de conjunt (*ensemble*) d'aquesta—. Rosenberg i Bouchard afirmen que el Principi de Selecció Natural és un principi causal perquè es basa a comparar les capacitats causals dels organismes per a sobreviure i reproduir-se (en el nivell individual).¹⁹ Aquest caràcter comparatiu s'observa en treballs teòrics com el de Wagner (2010), el qual elabora un test competitiu on dos genotips diferents competeixen entre ells, sent el guanyador aquell que aconsegueix ocupar més espai que l'altre, permetent la comparació entre diferents genotips i l'elaboració d'una escala de proporció de l'eficàcia.²⁰ Millstein (2006) replica que la selecció és causal però a nivell poblacional perquè, en necessitar almenys dos individus (la comparació de les capacitats dels organismes), la noció de Rosenberg i Bouchard col·lapsa en la noció poblacional. Aquesta visió no sembla allunyar-se de la usada pels biòlegs (veure LYNCH 2007 i RICE 2004).

¹⁹ Els autors defineixen, amb una redacció molt particular, aquest principi així: "PNS (x) (y) (E) [If x and y are competing organisms in generation n, and x is fitter than y in environment E, then probably, (there is some generation n', at which x has more descendants than y)]" (BOUCHARD i ROSENBERG 2004, 699).

²⁰ En realitat, aquesta elaboració teòrica no fa més que plasmar els experiments amb microorganismes realitzats des de fa dècades (veure ELENA i LENSKI 2003, i LUQUE 2016C).

La majoria dels filòsofs de la biologia que han entrat en la discussió s'han decantat per acceptar que la noció poblacional és necessària per a albirar el poder causal de la selecció i la deriva (STEPHENS 2004; REISMAN i FORBER 2005; MILLSTEIN 2006; SHAPIRO i SOBER 2007; ABRAMS 2007), encara que uns altres continuen defensant la noció causal a nivell d'individus (GLYMOUR 2006; GILDENHUYS 2011; BOURRAT 2019).²¹ Aquest debat, tanmateix, fa l'efecte de ser més un debat semàntic, això és, sobre termes i definicions, que un debat teòric substantiu. És clar que és necessari que existisca una població d'individus perquè es produïska l'evolució, ja que el que evolucionen són les poblacions i no els individus. Si definim com a procés poblacional aquell que requereix la presència de més d'un organisme, llavors podem denominar procés poblacional a la selecció o a la deriva. Però ha de quedar clar que tots dos processos, en contra del que defensa la VE, no són un element extra o *tertium quid* diferent dels organismes i la seua interacció amb l'ambient (HITCHCOCK i VELASCO 2014; LUQUE 2016A i 2016C).

Una possible crítica a la visió dinàmica és que no ha aconseguit articular una proposta causal unificada, desenvolupant els filòsofs de la biologia diferents concepcions causals en les quals integrar i articular la teoria evolutiva. Una de les més importants està basada en la noció intervencionista de la causalitat elaborada per Woodward (2003), la qual defensa que podem afirmar que C és una causa d'E si som capaços de manipular sistemàticament C perquè produïska un canvi en E. Vejam un exemple de causalitat *a la Woodward* en biologia evolutiva. Reisman i Forber (2005) analitzen l'experiment de Dobzhansky i Pavlovsky (1957) que va ser la primera demostració de l'efecte fundador. Dobzhansky i Pavlovsky van iniciar poblacions de laboratori de *Drosophila pseudoobscura* a partir d'una població en què una certa reordenació del tercer cromosoma (denominada PP) tenia una freqüència de 0.5. Es van establir dos tipus de poblacions, unes grans iniciades amb uns 5.000 individus i altres xicotetes amb 20 individus cadascuna. Al cap d'un any i mig (unes 18 generacions) els experimentadors van observar que la variació en les poblacions xicotetes era molt major que en les grans degut a aquest efecte fundador (una forma de deriva genètica), alterant el resultat de la selecció —la qual afavoria als heterozigòtics—. Per tant, en la línia de la concepció intervencionista de

²¹ Cal no confondre aquest argument amb el problema dels nivells o unitats de selecció. En aquests últims s'ha debatut específicament quins són els beneficiaris dels trets que afavoreix la selecció natural. Això es deu, en gran manera, a la postulació de l'existència d'adaptacions que afavoreixen, no als individus en particular, sinó al grup en el seu conjunt. Per contra, la discussió en el text principal fa referència a on establir la capacitat causal de la selecció. És a dir, tots els autors citats acceptarien un exemple clàssic de trets que afavoreixen als individus d'una població i no al grup (unitat de selecció: individus), però no estarien d'acord sobre on establir el poder causal de la selecció (nivell poblacional o nivell individual).

Woodward, la selecció i la deriva són causes del canvi evolutiu perquè alterant cadascun dels seus paràmetres (les eficàcies dels individus en un, la grandària poblacional en l'altre) s'obtenen sistemàticament canvis en les freqüències gèniques. La selecció i la deriva actuarien com a elements causals determinista i estocàstic, respectivament, com ocorre en molts sistemes físics. Així, si deixem caure una pedra pesada i un full de paper, la primera farà un recorregut simple i recte al terra mentre que la segona planejarà cap avall amb un recorregut complicat i erràtic. Encara que en tots dos casos tenim dues forces actuant, la força de la gravetat i la de fricció, el full de paper es veurà molt més afectada per la segona força mentre que la pedra patirà només xicotetes perturbacions (FUCHS 2013). De la mateixa manera, una població xicoteta es veurà més afectada per la deriva, mentre que una població gran es veurà més afectada per la selecció.

Les diferents interpretacions causals que han entrat en joc en aquest debat es poden englobar, en principi, en tres grans blocs:²² una visió intervencionista de la causalitat basada en Woodward (2003), que engloba diferents propostes com la interpretació de forces, la interpretació processal, etc. (REISMAN i FORBER 2005; SHAPIRO i SOBER 2007; MILLSTEIN 2006; CLATTERBURK 2015); una visió contrafàctica de la causalitat (GLENNAN 2009; HUNEMAN 2012); i una visió probabilística de la causalitat (RAZETO-BARRY i FRICK 2011; OTSUKA 2016). Aquesta varietat d'interpretacions causals podria ser considerada un punt feble en la visió causal de la teoria evolutiva. Tanmateix, això seria així només si la visió casual es veiés obligada alhora a salvar el suposat valor causal de cadascun dels factors que intervenen en la teoria apel·lant a nocions diferents de causa. Podria haver-hi pluralitat (i discussió entre filòsofs) però imaginem que qualsevol de les tres opcions sobre causalitat encaixara perfectament respecte als factors que intervenen en la teoria evolutiva. Llavors no hi hauria problema per als filòsofs de la biologia, però sí que n'hi hauria, en tot cas, per als filòsofs que discuteixen sobre la causalitat. Certament sembla desitjable que s'adopti una visió causal que englobe, si no totes, una gran part de les interpretacions causals (LUQUE 2016A i 2016B).

4.4 Deriva genètica: un procés sense direcció?

L'analogia de forces sembla trobar-se amb un gran problema en tractar amb la deriva genètica. Encara que podem atribuir una magnitud a la deriva genètica —la grandària poblacional N — no podem, en canvi, assignar-li una direcció. De fet, la deriva genètica per definició manca de direcció, és un

²² Per descomptat, en la literatura referent a la causalitat existeixen altres interpretacions causals no tractades en el debat. Veure Beebe *et al.* (2009) per a una introducció sistemàtica sobre causalitat.

procés no-direccional en tant que el canvi esperat en cada punt es zero (RICE 2004, 131). No obstant això, diversos autors han intentat, si no resoldre aquest problema, almenys matisar-lo prou per a continuar considerant la deriva com una força evolutiva en tant que constitueix un component causal essencial del fenomen evolutiu.

L'analogia amb la mecànica newtoniana s'ha posat en dubte a causa de les diferències entre la deriva i la resta de forces evolutives (LUQUE 2016A). Això es deu al fet que la deriva genètica posseeix diverses propietats diferents a la resta. La més important és que la deriva, per definició, no té direcció pel fet que les probabilitats que un gen o tret augmente o disminueixi en una població es cancel·len entre elles. Per tant, el canvi esperat és igual a zero, $E(\Delta\bar{z}) = 0$. Tanmateix, diversos autors han tractat de protegir l'estatus de força per a la deriva genètica: negant la seua suposada falta de direccionalitat (Stephens), estenent la noció de força (Filler) i cercant una força que també manque de direcció en física (Pence). Stephens (2004; 2010) postula que la deriva sí que té una direcció: condueix a les poblacions a l'homozigosi. Certament, un dels efectes d'una població sotmesa a la deriva genètica és que redueix la variabilitat al·lèlica present i, arribat un moment, es fixarà un tipus en la població. Filler (2009) per part seua, es mostra partidari de mantindre el parlar de forces (*force-talk*), però és conscient que un abús del terme pot buidar de contingut el mateix concepte de força. Així Filler postula dos criteris. El primer és que ha d'unificar fenòmens diversos, permetent-nos veure què tenen en comú aquests fenòmens. El segon és que ha de tindre una magnitud matemàtica específica precisa. Tots dos criteris els compleix amb escriu la deriva. Pence (2017) cerca un altre camí, en anar directament cap a la física per a trobar un fenomen que s'assembla prou a la deriva i que pugui ser considerat una força newtoniana. Pence afirma haver-ho trobat en el moviment brownià, el moviment fluctuant de les partícules en un líquid. El moviment brownià es modela estocàsticament mitjançant l'equació de Langevin, equivalent a la segona llei de Newton per a fenòmens estocàstics.

Encara que aquestes estratègies finalment tenen èxit (essencialment la de Pence), aquesta discussió passa per alt el punt crucial en el debat entre la VD i VE: l'estatus causal i l'estructura interna de la teoria evolutiva. Part de la discussió entre els VD i VE, lluny de millorar la nostra comprensió de l'estructura i causes de la teoria evolutiva, s'ha centrat algunes vegades en aspectes poc importants o tangencials —per exemple, la quantitat de pàgines dedicades a discutir sobre llançaments de monedes (WALSH *et al.* 2002; STEPHENS 2004 i 2010; PIGLIUCCI i KAPLAN 2006; WALSH 2007; OTSUKA *et al.* 2011; SARKAR 2011; CLATTERBUCK *et al.* 2013) i no sobre exemples biològics— però no aclarint necessàriament els elements fonamentals de la discussió: els models que

conformen la teoria evolutiva, la història de com es van construir aquests models, i l'ús que fan d'ells els mateixos investigadors. Seguint aquests elements fonamentals (els quals he delinat parcialment en el present article) podrem millorar la nostra comprensió de l'estructura i causes de la teoria evolutiva (tornaré sobre aquest punt en les Conclusions).

4.5 La deriva com a llei d'estat zero

Hi ha una altra possibilitat per a la deriva. El caràcter especial d'aquesta —sent l'única de les anomenades forces evolutives que manca de direcció i que és estocàstica, mentre que la resta tenen direcció i són deterministes— pot ser perquè que no estiguem davant d'una força més, sinó que estiguem davant el “estat per defecte” o, en termes newtonians, la llei de força zero de qualsevol sistema evolutiu. Alguns autors han explorat aquesta possibilitat. McShea i Brandon han defensat aquest l'estatus especial per a la deriva en detriment de la llei Hardy-Weinberg (MCSHEA i BRANDON 2010; BRANDON 2006, 2010). Ells postulen que la deriva, lluny de ser una força especial que s'introdueix en la població, és l'estat per defecte de les poblacions i per tant una llei de força zero, de la mateixa manera que la inèrcia és l'estat per defecte dels cossos en la mecànica newtoniana. En una línia d'argumentació similar, Sarkar (2011) situa a la deriva com l'estat per defecte de les poblacions. Sarkar elabora un model haploide amb una població tancada (no té migració) i que segueix les lleis mendelianes, en el qual només es tindran en compte la selecció, la mutació i la deriva. Sarkar assenyala que en cap moment s'ha esmentat la deriva en el model i, no obstant això, apareix en la seua elaboració a través de la grandària poblacional quan aquesta és finita. D'aquesta manera, la selecció i la mutació expressades en el model serien causes del canvi evolutiu. La deriva, en canvi, no seria una causa del canvi evolutiu, sinó una part constitutiva del sistema. Res causa que hi haja canvi on es dona un procés de mostreig constant en les poblacions finites, és el seu estat per defecte i d'ell sorgeix el caràcter estocàstic dels models en poblacions finites.

En Luque (2016B) done diferents raons de perquè la deriva no pot funcionar com una llei d'estat zero. La deriva posseeix unes característiques que no corresponen amb les característiques d'una llei d'aquest tipus. Aquestes lleis són intrínsecament idealitzacions que mai no poden realitzar-se de forma efectiva en sistemes reals i, a més, estableixen què és el que requereix explicació en el sistema. Així, la llei de la inèrcia estipula que és el canvi de velocitat d'un objecte el que requereix explicació. A més a més, com aquesta llei és una idealització (mai cap cos real pot estar en un sistema newtonià lliure de la influència de cap força), qualsevol cos que mantinga la seua velocitat constant

serà degut a diferents forces que s'anul·len entre elles. La deriva, en canvi, és un procés *real* al que estan sotmesos les poblacions, i els biòlegs li atorguen poder causal i explicatiu. En canvi, propose que el comportament per defecte dels sistemes evolutius és el que he anomenat el Principi d'Estasis (*Principle of Stasis*) que postula que si no hi ha cap causa evolutiva actuant en el sistema, aquest es mantindrà sense canvis (veure LUQUE 2016B per a més detalls).

5. CONCLUSIONS: UNA PROPOSTA DE RESOLUCIÓ DEL DEBAT

El present article ha fet una anàlisi del debat produït en els últims vint anys respecte a l'estructura causal de la teoria evolutiva. Originalment es va entendre com una teoria de forces. Amb l'objectiu de posar en relleu fins a on aquesta interpretació podia ser correcta, es van plantejar diferents crítiques i possibles solucions. Aquesta estratègia ens permet delimitar-la i traçar un perfil precís del seu abast. Crec que, entesa de manera correcta, la interpretació de forces ens ofereix una imatge correcta de l'estructura de la teoria evolutiva. Al meu entendre, l'estructura de la teoria evolutiva es compon de tres pilars fonamentals:

Llei d'estat zero. El Principi d'Estasis.

Llei conseqüencial. L'equació de Price.

Lleis font. Aquestes serien les diferents forces evolutives i les seues corresponents formulacions matemàtiques: selecció natural, deriva, mutació, migració, etc.

Com podem veure, tenim la mateixa estructura delineada per Sober. Tanmateix, les lleis d'estat zero i conseqüencial són diferents a les establertes originalment per Sober en 1984. Cadascuna compleix el seu paper establert. El Principi d'Estasis ens diu com es comporta el sistema si cap força o factor evolutiu actua sobre les poblacions. Aquestes forces evolutives estan englobades en les lleis font. Respecte a la llei conseqüencial, encara que els models de la genètica de poblacions han realitzat tradicionalment aquest paper, considere que l'equació de Price és el candidat idoni per al paper de llei conseqüencial de la teoria evolutiva. L'equació de Price unifica els diferents models simplificats elaborats al llarg del segle XX i anima als investigadors a trobar noves formes de la mateixa per a resoldre nous problemes –de manera similar a la funció que fa la segona llei de Newton en la mecànica clàssica (veure LUQUE 2017 per a més detalls). Un altre aspecte comú important és que totes dues lleis conseqüencials permeten arribar a la seua pròpia llei d'estat zero. En mecànica clàssica, si no existeix cap força, $F = 0$, llavors el cos es manté a velocitat constant, és a dir, segueix la llei de la inèrcia. En la teoria evolutiva, si no existeix cap força evolutiva, $Cov(w, z) = 0$ i $E(w\Delta z) = 0$, llavors no hi haurà canvi en les freqüèn-

cies gèniques (genotípiques o qualsevol altre tret d'interés). És a dir, $\bar{w}\Delta\bar{z} = 0$, seguint el Principi d'Estasis. Al seu torn, aquestes lleis conseqüencials permeten incorporar la interacció de les forces, mostrant fins i tot casos de simulacions d'estats zero quan es cancel·len entre elles. Seria el cas en mecànica clàssica de dues forces, $F1$ i $F2$, que actuen sobre el mateix cos amb igual magnitud i direcció però sentit contrari. És a dir: $F1 + (-F2) = 0$. En sistemes evolutius ho trobaríem, per exemple, amb la interacció entre la mutació, $E(w\Delta z)$, i la selecció $Cov(w, z)$, com en el cas d'equilibri mutació-selecció, on $Cov(w, z) + (-E(w\Delta z)) = 0$.

Malgrat tota la discussió desenvolupada sobre si les forces evolutives eren comparables a les forces newtonianes, l'estructura apuntada anteriorment indica que la majoria de debats han fallat en el punt central de la discussió. Aquesta deuria d'haver-se centrat en l'estructura interna de la teoria evolutiva (i de la mecànica newtoniana) i no en els elements externs d'ambdues. Per dir-ho en termes carnapians, no estem avaluant les similituds entre la teoria evolutiva i la newtoniana respecte a les seues concepcions "externes" sobre com el món és. Ans al contrari, estem avaluant les virtuts de cada marc conceptual considerant els seus objectius epistèmics i explanatoris "interns". Així, els avantatges d'aquesta estructura causal per a la teoria evolutiva són múltiples.²³ En primer lloc, elimina els problemes d'estendre el concepte de força més enllà de la mecànica clàssica, establint que una força evolutiva serà allò que afecte a alguns dels termes en l'equació de Price (com, de fet, fa la deriva genètica). En segon lloc, aquest marc causal és aplicable tant a la teoria evolutiva com a la mecànica newtoniana, i ens mostra quines són les seues semblances estructurals i la utilitat de la interpretació de forces. L'analogia newtoniana és il·luminadora perquè revela l'estructura interna de la teoria evolutiva i les causes que hi postula. La teoria està construïda a partir d'una llei d'estat zero que estipula un comportament per defecte i introdueix factors (les anomenades forces evolutives) que alteren aquest comportament. Aquesta és la raó de per què es va formular l'analogia de forces en el passat i per què continua hui en dia en la literatura biològica.

AGRAÏMENTS: Agraesc a un revisor anònim els seus comentaris, que milloraren el present text. Aquest treball ha sigut finançat pel Ministerio de Ciencia, Innovación y Universidades mitjançant un contracte Juan de la Cierva-formación (FJC2018-038661-I), i els projectes FFI2016-76799-P i FFI2017-89639-P.

²³ Per a més detalls sobre les possibilitats d'aquesta nova versió de la interpretació de forces, veure Luque (2018).

APÈNDIX

A1. Model matemàtic de Sewall Wright

La fórmula de la selecció independent de freqüència desenvolupada per Wright és la següent:

$$\Delta p = \frac{p(1-p)}{2\bar{w}} \frac{d\bar{w}}{dp}$$

on p és la freqüència d'un al·lel, \bar{w} és l'eficàcia mitjana poblacional, i $\frac{d\bar{w}}{dp}$ és el pendent de l'eficàcia mitjana. La direcció de l'evolució (és a dir, si p augmenta o disminueix) és proporcional a (per tant està determinat per) el pendent de la funció de l'eficàcia mitjana, canviant sempre de tal manera que l'eficàcia poblacional mitjana augmenta. Com $\frac{d\bar{w}}{dp}$ és igual a $\frac{d \ln(\bar{w})}{dp}$ i $\ln(\bar{w}) = \bar{m}$ és la taxa de creixement instantani per capita de la població, la selecció natural fa que la població *escale* cap al pic adaptatiu més pròxim —si existeixen diferents pics, la selecció espenta a la població a l'òptim més pròxim, que pot ser un òptim local i no necessàriament l'òptim global— escalant la població cap a punts cada vegada més alts en la corba de (GILLESPIE 2004; RICE 2004) com mostra la Figura 4.

A2. Model d'equilibri mutació-selecció

Suposem que q és la freqüència de la classe al·lèlica B i $p = 1 - q$ és la freqüència del classe al·lèlica A , per la qual cosa la mutació porta al següent canvi evolutiu:

$$\Delta q = p\mu$$

on μ és la taxa de canvi que cada còpia d' A se convertisca en B . Ara imaginem que hi ha selecció en contra de l'al·lel, per tant la fitness dels tres genotips AA , AB i BB són 1 , $1 - hs$, i $1 - s$ respectivament. Llavors, l'excés de fitness en mitjana de l'al·lel B (suposant aparellament a l'atzar) és:

$$B_a = p(1 - hs - \bar{w}) + q(1 - s - \bar{w})$$

Si la població està prop de l'equilibri selectiu, llavors $p \approx 1$, per tant $\Delta q = \mu$. Si suposem a més que, en estar prop de l'equilibri selectiu, quasi tots els individus de la població són homozigots AA , llavors la fitness mitjana $\bar{w} \approx 1$. Després de simplificar, obtenim que l'equilibri entre mutació i selecció ve donat per

$$\hat{q} = \frac{\mu}{hs + \hat{q}s(1-h)}$$

Encara que la seua forma més coneguda ve donada per a dominància completa ($h = 0$), quedant l'equació per a un al·lel recessiu deleteri autosòmic en equilibri (Templeton 2006)

$$\hat{q}^2 = \frac{\mu}{s} \rightarrow \hat{q} = \sqrt{\frac{\mu}{s}}$$

Llavors, la freqüència de l'al·lel deleteri es mantindrà en equilibri en la població perquè la selecció serà incapaç d'eliminar-ho en estar “amagat” entre els genotips heterozigòtics. Aquest model explica la gran quantitat de mutacions rares de les quals són portadores molts cromosomes en poblacions de diverses espècies (*Drosophila*, humans, etc.), les quals redueixen l'eficàcia dels heterozigòtics i són fins i tot letals per a l'homozigòtic portador de l'al·lel deleteri (FUTUYMA 2013).

REFERÈNCIES BIBLIOGRÀFIQUES

- ABRAMS, M. 2007, "How Do Natural Selection and Random Drift Interact?", *Philosophy of Science* 74: 666-79.
- BATESON, W. i SAUNDERS, E. R. 1902, "The Facts of Heredity in the Light of Mendel's Discovery", *Reports to the Evolution Committee of the Royal Society* I: 125-60.
- BEEBEE, H.; HITCHCOCK, Ch. i MENZIES, P. (ed.) 2009, *The Oxford Handbook of Causation*, Oxford: Oxford University Press.
- BOUCHARD, F. i ROSENBERG, A. 2004, "Fitness, Probability, and the Principles of Natural Selection", *British Journal for the Philosophy of Science* 55: 693-712.
- BOURRAT, P. 2019, "Evolution is about populations, but its causes are about individuals", *Biological Theory* 14(4): 254-66.
- BRANDON, R. 2006, "The principle of drift: biology's first law", *Journal of Philosophy* 103(7): 319-35.
- BRANDON, R. 2010, "A non-Newtonian model of evolution: the ZFEL view", *Philosophy of Science* 77(5): 702-15.
- BRANDON, R. i RAMSEY, G. 2007, "What's Wrong with the Emergentist Statistical Interpretation of Natural Selection and Random Drift?", D. HULL i M. RUSE (ed.), *Cambridge Companion to the Philosophy of Biology*. Cambridge: Cambridge University Press, 66-84.
- CAPONI, G. 2004, "La navaja de Darwin", *Ludus Vitalis* 12(22): 9-38.
- CLATTERBUCK, H. 2015, "Drift Beyond Wright-Fisher", *Synthese* 192: 3487-507
- CLATTERBUCK, H.; SOBER, E. i LEWONTIN, R. 2013, "Selection never dominates drift (nor vice versa)", *Biology and Philosophy* 29: 1-16.
- CORBEN, H. C. i STEHLE, P. 1994 (2a edició), *Classical Mechanics*, Nova York: Dover.
- CROW, J. i KIMURA, M. 1970, *An introduction to population genetics theory*. Minneapolis: Burgess Publishing Company.
- DARWIN, CH. 1859, *On the origin of species*, Londres: John Murray.
- DE VRIES, H. 1904, *Species and varieties: Their origin by mutation*, Chicago: The Open Court Publishing Company.
- DOBZHANSKY, T. 1951, *Genetic and the Origin of Species (third edition)*, Nova York: Columbia University Press.
- DOBZHANSKY, T. i PAVLOVSKY, O. 1957, "An Experimental Study of the Interaction between Genetic Drift and Natural Selection", *Evolution* 11: 311-9.
- ELENA, S. F. i LENSKI, R. E. 2003, "Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation", *Nature Reviews Genetics* 4: 457-69.
- FAIRBAIN, D. J. i REEVE, J. P. 2001, "Natural Selection", CH. FOX, D. ROFF i D. FAIRBAIN (ed.), *Evolutionary Ecology. Concepts and Case Studies*, Oxford: Oxford University Press, 29-43.

- FILLER, J. 2009, "Newtonian forces and evolutionary biology: a problem and solution for extending the force interpretation", *Philosophy of Science* 76: 774-83.
- FISHER, R. 1930, *The genetical theory of natural selection*, Londres: Clarendon Press.
- FONTDEVILA, A. i MOYA, A. 1999, *Introducción a la genética de poblaciones*, Madrid: Síntesis.
- FORBER, P. i REISMAN, K. 2007, "Can there be stochastic evolutionary causes?", *Philosophy of Science* 74(5): 616-27
- FREEMAN, S. i HERRON, C. 2007 (4a edició), *Evolutionary Analysis*, Londres: Pearson.
- FUCHS, A. 2013, *Nonlinear Dynamics in Complex Systems*, Berlín: Springer.
- FUTUYMA, D. J. 2013 (3a edició), *Evolution*. Sunderland: Sinauer.
- GILDENHUYS, P. 2011, "Righteous modeling: the competence of classical population genetics", *Biology and Philosophy* 26(6): 813-35.
- GILLESPIE, J. 2004 (2a edició), *Population Genetics: A Concise Guide*, Baltimore: The John Hopkins University Press.
- GLYMOUR, B. 2006, "Wayward modeling: population genetics and natural selection", *Philosophy of Science* 73: 369-89.
- GOULD, S. J. 2002, *The structure of evolutionary theory*, Nova York: Belknap Press.
- HALDANE, J. B. S. 1932, *The Causes of Evolution*, Londres: Longmans, Green and Company.
- HARTL, D. i CLARK, A. 1997 (3a edició), *Principles of population genetics*, Sunderland: Sinauer.
- HITCHCOCK, H. i VELASCO, J. 2014, "Evolutionary and Newtonian forces", *Ergo* 1(2): 39-77.
- HUNEMAN, P. 2012, "Natural selection: A case for the counterfactual approach", *Erkenntnis* 76(2): 171-94.
- LANDE, R. i ARNOLD, S. 1983, "The measurement of selection on correlated characters", *Evolution* 37: 1210-26.
- LEWONTIN, R. 1970, "The units of selection", *Annual Review of Ecology and Systematics* 1: 1-18.
- LUQUE, V. J. 2016A, "Drift and evolutionary forces: scrutinizing the Newtonian analogy", *Theoria* 31(3): 397-411.
- LUQUE, V. J. 2016B, "The Principle of Stasis. Why drift is not a Zero-Cause Law", *Studies in History and Philosophy of Biological and Biomedical Sciences* 57: 71-9.
- LUQUE, V. J. 2016C, "From toys to games: overcoming the view of natural selection as a filter", *Kairos* 17(1): 1-24.
- LUQUE, V. J. 2017, "One equation to rule them all. A philosophical analysis of the Price equation", *Biology and Philosophy* 32(1): 97-125.
- LUQUE, V. J. 2018, "Is there room for a new foundation for the force interpretation?", C. SABORIDO, S. OMS i J. GONZÁLEZ DE PRADO, *Actas del IX Con-*

- greso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España, 353-7.
- LYNCH, M. 2007, "The Frailty of Adaptive Hypotheses for the Origins of Organismal Complexity", *PNAS* 104: 8597-604.
- MARSHALL, J. A. R. 2009, "The donation game with roles played between relatives", *Journal of Theoretical Biology* 260: 386-91.
- MATTHEN, M. i ARIEW A. 2002, "Two ways of thinking about fitness and natural selection", *Journal of Philosophy* 99(2): 55-83.
- MATTHEN, M. i ARIEW, A. 2005, "How to understand causal relations in natural selection: reply to Rosenberg and Bouchard", *Biology and Philosophy* 20: 355-64.
- MATTHEN, M. i ARIEW, A. 2009, "Selection and causation", *Philosophy of Science* 76: 201-24.
- MAUDLIN, T. 2004, "Causation, counterfactuals, and the third factor", J. D. COLLINS, N. HALL i L. A. PAUL (ed.), *Causation and counterfactuals*, Cambridge: The MIT Press, 419-43.
- MCGHEE, G. 2007, *The Geometry of Evolution*, Cambridge: Cambridge University Press.
- MC SHEA, D. i BRANDON, R. 2010, *Biology's first law: the tendency for diversity and complexity to increase in evolutionary systems*, Chicago: University of Chicago Press.
- MILLSTEIN, R. 2002, "Are random drift and natural selection conceptually distinct?", *Biology and Philosophy* 17(1): 33-53
- MILLSTEIN, R. 2006, "Natural selection as a population-level causal process", *British Journal for Philosophy of Science* 57(4): 627-53
- NEWTON, I. 1846 [1687], *The Mathematical Principles of Natural Philosophy*, Nova York: Daniel Adee.
- OKASHA, S. 2006, *Evolution and the levels of selection*, Oxford: Oxford University Press
- OKASHA, S. 2008, "Fisher's Fundamental Theorem of Natural Selection—A Philosophical Analysis", *British Journal for the Philosophy of Science* 59: 319-51.
- ORR, H. A. 2009, "Fitness and its role in evolutionary genetics", *Nature Reviews* 10: 531-9.
- OTSUKA, J. 2016, "Causal foundations of evolutionary genetics", *British Journal for Philosophy of Science* 67: 247-69
- OTSUKA, J.; TURNER, T.; ALLEN, C. i LLOYD, E. 2011, "Why the causal view of fitness survives", *Philosophy of Science* 78(2): 209-24
- PENCE, CH. 2017, "Is Genetic Drift a Force?", *Synthese* 194(6): 1967-88.
- PIGLIUCCI, M. i KAPLAN, J. 2006, *Making sense of evolution: the conceptual foundations of evolutionary theory*, Chicago: University of Chicago Press.

- PLUTYNSKI, A. 2007, "Drift: A historical and conceptual overview", *Biological Theory* 2(2): 156-67.
- RAZETO-BARRY, P. i FRICK, R. 2011, "Probabilistic causation and the explanatory role of natural selection", *Studies in History and Philosophy of Biological and Biomedical Sciences* 42: 344-55.
- REISMAN, K. i FORBER, P. 2005, "Manipulation and the Causes of Evolution", *Philosophy of Science* 72: 1113-23.
- RICE, S. H. 2004, *Evolutionary Theory: Mathematical and Conceptual Foundations*, Sunderland: Sinauer.
- ROSENBERG, A. i BOUCHARD, F. 2005, "Matthen and Ariew's Obituary for Fitness: Reports of Its Demise Have Been Greatly Exaggerated", *Biology and Philosophy* 20: 343-53.
- ROUGHGARDEN, J. 1996, *Theory of Population Genetics and Evolutionary Ecology: An Introduction*, Upper Saddle River: Prentice Hall.
- ROUGHGARDEN, J. 1998, *Primer of Ecological Theory*, Upper Saddle River: Prentice Hall.
- RUSE, M. 1973, *The Philosophy of Biology*, Londres: Hutchinson
- SARKAR, S. 2011, "Drift and the causes of evolution", P. MCKAY, F. RUSSO i J. WILLIAMSON (ed.), *Causality in the Sciences*, Oxford: Oxford University Press, 444-69.
- SHAPIRO, L. i SOBER, E. 2007, "Epiphenomenalism: the do's and the don'ts", G. WOLTERS i P. MACHAMER (ed.), *Thinking about causes: from Greek philosophy to modern physics*, Pittsburgh: University of Pittsburgh Press, 235-64.
- SKIPPER, R. i DIETRICH, M. 2012, "Sewall Wright's Adaptive Landscape: Philosophical Reflections on Heuristic Value", E. SVENSSON i R. A. CALSBEEK (ed.), *The Adaptive Landscape in Evolutionary Biology*, Nova York: Oxford University Press, 16-25.
- SOBER, E. 1984, *The Nature of Selection*, Cambridge: MIT Press.
- STEPHENS, CH. 2004, "Selection, drift, and the 'forces' of evolution", *Philosophy of Science* 71(4): 550-70.
- STEPHENS, CH. 2010, "Forces and Causes in Evolutionary Theory", *Philosophy of Science* 77(5): 716-27.
- STOLTZFUS, A. i CABLE, K. 2014, "Mendelian-Mutationism: The Forgotten Evolutionary Synthesis", *Journal of the History of Biology* 47: 501-46.
- TEMPLETON, A. R. 2006, *Population Genetics and Microevolutionary Theory*, Nova Jersey: Wiley.
- TENAILLON, O. 2014, "The Utility of Fisher's Geometric Model in Evolutionary Genetics", *The Annual Review of Ecology, Evolution, and Systematics* 45: 179-201.
- WAGNER, G. P. 2010, "The measurement theory of fitness", *Evolution* 64(5): 1358-76.

- WALSH, D. 2007, "The pomp of superfluous causes: the interpretation of evolutionary theory", *Philosophy of Science* 74: 281-303.
- WALSH, D.; LEWENS, T. i ARIEW, A. 2002, "The trials of life: natural selection and random drift", *Philosophy of Science* 69(3): 429-46.
- WALSH D. M.; ARIEW, A. i MATTHEN, M. 2017, "Four Pillars of Statisticalism", *Philosophy and Theory in Biology* 9 (20170609).
- WILLIAMS, G. C. 1992, *Natural selection: domains, levels, challenges*, Oxford: Oxford University Press.
- WOODWARD, J. 2003, *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.
- WRIGHT, S. 1931, "Evolution in Mendelian populations", *Genetics* 16: 97-159.

Documents

ALBERTO OYA

IFILNOVA – Instituto de Filosofia da Nova
Arg Lab – Lisbon Mind and Reasoning Group
Universidade Nova de Lisboa, Portugal

Presentación.

Las definiciones persuasivas según Charles L. Stevenson

EL FILÓSOFO NORTEAMERICANO CHARLES LESLIE STEVENSON (1908, Cincinnati, Ohio – 1979, Bennington, Vermont) es conocido principalmente por formular y defender una concepción emotivista del lenguaje ético y, más concretamente, por ofrecer una explicación del desacuerdo ético compatible con la tesis básica emotivista que los juicios éticos no describen o representan hechos del mundo sino que *expresan* los sentimientos de aprobación o desaprobación de quien emite dichos enunciados. La formulación más sistemática de su análisis del desacuerdo ético se encuentra en la obra *Ethics and Language* (New Haven, Yale University Press, 1944), aunque su origen puede rastrearse en sus ensayos publicados en la década de los años treinta, muchos de ellos incluidos posteriormente en la compilación *Facts and Values: Studies in Ethical Analysis* (New Haven, Yale University Press, 1963).

El análisis del desacuerdo ético es, obviamente, un tema filosóficamente relevante en sí mismo, pero se torna un tema crítico para aquellos que, como Stevenson, defienden una concepción emotivista del lenguaje ético. La tesis principal del emotivismo metaético es, como ya dije, la afirmación que los juicios éticos no describen el mundo sino que expresan los sentimientos de aprobación o desaprobación de quien emite dichos enunciados. Así, por ejemplo, de acuerdo con el emotivismo, cuando emitimos juicios éticos del tipo “Robar es (éticamente) incorrecto”, no estamos *describiendo* ninguna *propiedad moral* predicable del acto de robar que lo haga éticamente inco-

recto, sino que estamos, simplemente, expresando nuestra actitud personal de desaprobación hacia el acto de robar. Esto implica, entre otras cosas, que el desacuerdo ético no involucra ninguna diferencia respecto a si un determinado acto *es*, de hecho, éticamente correcto o incorrecto, sino solamente la expresión de preferencias personales distintas. Es decir, y volviendo al ejemplo anterior, si decimos “Robar es (éticamente) incorrecto” y alguien nos responde que “Robar es (éticamente) correcto”, quien nos responde así no está realmente negando la afirmación que el acto de robar *es* (éticamente) incorrecto sino que está solamente expresando su actitud personal de aprobación respecto al acto de robar.

El problema para el emotivismo es que no diríamos que nuestras discusiones éticas son una mera contraposición de preferencias personales, pues éstas parecen centradas en resolver si un acto dado *es* (éticamente) correcto o incorrecto. De hecho, parece claro que cuando entramos en una discusión estamos asumiendo, de una manera más o menos implícita, que podemos convencer a nuestro oponente y que, en caso de que lo convencemos, éste dejará de lado su posicionamiento inicial y aceptará el nuestro. Es difícil entender por qué entraríamos en ninguna discusión, ya sea ética o de cualquier otro tipo, si creyéramos que no podemos convencer a nuestro oponente o si pensáramos que, aún convenciendo a nuestro oponente, éste no abandonaría su planteamiento inicial. Es decir, no parece que discutamos con el único propósito de expresar nuestras preferencias personales sino con el objetivo primario de negar las opiniones de nuestro oponente y convencerle de las nuestras. El reto para el emotivista es, pues, ofrecer un análisis del desacuerdo ético que explique por qué *parece* que estemos contradiciéndonos si *en realidad* sólo estamos expresando distintas preferencias personales.

La explicación del desacuerdo ético ofrecida por Stevenson pasa por distinguir entre dos tipos de significado: el llamado significado emotivo y el denominado significado conceptual. El primero refiere a la propiedad disposicional de un término a generar un determinado rango de emociones en quienes usan o reciben dicho término, mientras que el segundo refiere a la propiedad disposicional de un término a generar procesos mentales cognitivos como puede ser la formación de creencias. Es importante mencionar que estos dos tipos de significado, aunque diferenciables entre sí, no son, de acuerdo con Stevenson, mutuamente excluyentes y pueden darse ambos al mismo tiempo en un mismo término.

La distinción entre significado emotivo y significado conceptual permite a Stevenson distinguir dos tipos de desacuerdo ético: el desacuerdo en interés y el desacuerdo en creencia. El desacuerdo en interés refiere a una contraposición de intereses que no pueden ser satisfechos al mismo tiempo,

mientras que el desacuerdo en creencia refiere a una contraposición entre creencias que no pueden ser ambas verdad al mismo tiempo. Es decir, el desacuerdo en creencia es un desacuerdo acerca del valor de verdad de una descripción de como es de hecho el mundo, mientras que el desacuerdo en interés es un desacuerdo entre distintas formas de valorar el mundo.

Es de una importancia fundamental identificar en qué tipo de desacuerdo estamos inmersos si queremos encontrar un método adecuado de resolverlo. El desacuerdo ético, dice Stevenson, es primariamente un desacuerdo en interés, pero dado que, de acuerdo con Stevenson, los intereses y las creencias pueden influenciarse mutuamente, hay ocasiones en que el desacuerdo en interés está basado en un desacuerdo en creencia. El método empírico es útil para resolver el desacuerdo ético sólo en caso de que éste se base en un desacuerdo en creencia. Esto puede suceder de diversos modos, siendo el principal el de mostrar que un determinado acto tiene alguna consecuencia fáctica hacia la que tanto nosotros como nuestro oponente tenemos el mismo interés. En cambio, si el desacuerdo ético es puramente un desacuerdo en interés y no depende, por tanto, de ningún desacuerdo en creencia, el método empírico carece de utilidad. Esto no quiere decir, sin embargo, que el desacuerdo en interés no pueda ser resuelto por otros medios. Uno de estos medios es el uso de definiciones persuasivas, mediante las que se otorga un nuevo significado conceptual a un término ya en uso manteniendo su significado emotivo original.

El ensayo “Definiciones persuasivas”, publicado originalmente en 1938 y que se ofrece aquí por vez primera en lengua castellana, se enmarca, pues, en esta explicación más general del desacuerdo ético y la distinción entre significado emotivo y significado conceptual —lo que Stevenson en este ensayo denomina “la correlación entre terminología e intereses”—. Con todo, es importante enfatizar que las definiciones persuasivas no son exclusivas del ámbito ético y, sobre todo, que si bien el análisis que ofrece Stevenson de las definiciones persuasivas puede incrementar la coherencia de la tesis emotivista, su análisis es independiente tanto de la aceptación de la tesis emotivista como de su teoría disposicional del significado. El análisis de Stevenson *sólo* requiere de aceptar que la carga evaluativa que acompaña ciertos términos puede resistir un cambio en su significado fáctico o proposicional. En cualquier caso, la importancia fundamental del ensayo de Stevenson se encuentra en que éste es el primero en señalar, con absoluta claridad, el fenómeno de las definiciones persuasivas. Es por ello que, aún si por alguna razón encontráramos el análisis de Stevenson inadecuado o incompleto, su ensayo sigue siendo, a día de hoy, el mejor punto de partida para el estudio de las definiciones persuasivas.

Nota sobre la edición

Para la traducción de “Persuasive Definitions” se ha usado el texto recopilado en *Facts and Values: Studies in Ethical Analysis* (New Haven, Yale University Press, 1963, 32-54). Esta edición incluye muy ligeras modificaciones por parte del propio autor respecto al texto original publicado en el número 47 de la revista *Mind* (1938).

Se incluyen dos tipos de notas a pie de página. Las que siguen la numeración arábica son originales de Stevenson; las que van marcadas con asterisco son del traductor.

CHARLES LESLIE STEVENSON

Definiciones persuasivas

Traducción de Alberto Oya

I

UNA DEFINICIÓN “PERSUASIVA” es aquella que da un nuevo significado conceptual a una palabra conocida sin cambiar sustancialmente su significado emotivo, y que se usa con el propósito consciente o inconsciente de cambiar, por este medio, la dirección de los intereses de las personas.^{1*}

El objetivo de este artículo es mostrar que las definiciones persuasivas son a menudo utilizadas en filosofía y que el error generalizado de no reconocerlas como tales —es decir: la tentación de considerarlas como definiciones que simplemente abrevian o que analizan conceptos comunes— ha llevado a importantes confusiones filosóficas.

¹ En este ensayo, como en el ensayo “The Emotive Meaning of Ethical Terms”, el término “interés” tiene el sentido dado por R. B. Perry, que en otros lugares de la compilación *Facts and Values* se expresa con el término “actitud”. Véase la nota número 2 de “The Emotive Meaning of Ethical Terms”.

* *Nota del traductor.* “The Emotive Meaning of Ethical Terms” [“El significado emotivo de los términos éticos”] fue publicado originalmente en el número 46 de la revista *Mind* (1937) y fue reimpresso en la compilación *Facts and Values: Studies in Ethical Analysis*, publicada en 1963 junto a otros ensayos de Stevenson, incluyendo el presente artículo sobre definiciones persuasivas. En “The Emotive Meaning of Ethical Terms” encontramos la formulación más temprana de su concepción emotivista del lenguaje ético, que años más tarde expone de manera sistemática en su *Ethics and Language* (1944). La nota a que hace referencia Stevenson dice así: “En su *General Theory of Value* (Nueva York, 1926), Perry usó ‘interés’ para referir a cualquier tipo de aprobación o desaprobación, o a cualquier tipo de disposición a estar a favor o en contra de algo. Y usó ‘teoría’ donde podría haber usado ‘definición propuesta’ o ‘análisis de un significado de sentido común’. En la mayoría de mis ensayos contenidos en *Facts and Values*, el término ‘interés’ da lugar sistemáticamente al término ‘actitud’. El propósito de este cambio fue sólo el de ofrecer una terminología más transparente —no se pretendía repudiar la concepción de interés de Perry” (*Facts and Values: Studies in Ethical Analysis*. New Haven, Yale University Press, 1963, 12)—.

Antes de considerar ejemplos filosóficos, será útil considerar algunos ejemplos más simples, que nos servirán para clarificar qué son las definiciones persuasivas.

Como primer ejemplo, consideremos una definición de la palabra “cultura”. Conviene que definamos de manera completamente ficticia los hábitos lingüísticos de las personas a quienes se dirige esta definición, pues ello librará nuestra situación real de complicaciones irrelevantes. Supongamos, pues, una comunidad hipotética en la que “cultura” comenzó teniendo un significado casi exclusivamente conceptual. Esbozcamos el desarrollo de su significado emotivo, mostremos por qué el significado emotivo llevó a algunas personas a redefinir la palabra, y examinemos la manera en que esta redefinición consiguió su propósito.

Hubo una vez una comunidad en la que “culto” significaba *ampliamente leído y versado en las artes*.

Con el paso del tiempo, estas cualidades tomaron más importancia. Si un hombre quería hacerle un cumplido a otro, hablaría extensamente de su cultura. Se tornó antinatural usar “cultura” en un tono de voz que no fuera laudatorio. Aquellos que carecían de cultura usaban la palabra con reverencia, y aquellos que la poseían usaban la palabra con satisfacción personal o quizás con una delicada modestia. De este modo la palabra adquirió un fuerte significado emotivo. Despertaba sentimientos no sólo por su significado conceptual sino de una forma más directa, por sí misma; pues despertaba el recuerdo de los gestos, las sonrisas y el tono de voz que tan habitualmente le acompañaban. Un orador, por ejemplo, no era nunca presentado como “un hombre ampliamente leído y versado en las artes”, sino que era descrito como “un hombre de cultura”. Esta última frase no se diferenciaba de la primera en su significado conceptual, pero era más adecuada para despertar en el público una actitud favorable.

Conforme el significado emotivo se hizo *más pronunciado, el significado conceptual se tornó más vago. Esto era inevitable, pues el significado emotivo hizo que la palabra se tornara adecuada para su uso metafórico*. Con frecuencia se llamó cultos a hombres que no eran cultos en el sentido literal, especialmente cuando éstos eran admirados por tener *algunas* de las cualidades definitorias de “cultura”. Al principio la gente distinguía fácilmente estos cumplidos metafóricos de los enunciados literales; pero a medida que las metáforas se hicieron más frecuentes, la distinción se tornó menos clara. La gente no estaba del todo segura si una persona *debía* ser versada en las artes con tal de ser culto en el sentido literal. Otro tipo de conocimiento quizás podría servir como sustituto.

Supongamos ahora que hubo un miembro de la comunidad que no tuvo en estima la mera lectura o el mero conocimiento de las artes en sí mismos, sino que los valoró en la medida en que éstos servían para desarrollar la sensibilidad

imaginativa. Sentía que no siempre eran un medio adecuado para conseguir tal fin, y que en ningún caso eran el único medio. Constantemente lamentaba que tales procedimientos mecánicos, como la lectura o el visitar museos, debieran recibir elogios de manera automática, y que la sensibilidad imaginativa apenas se tomara en cuenta. Es por esta razón que procedió a dar un nuevo significado a “cultura”. “Sé”, insistía él, “que tal o cual persona es ampliamente leída y versada en las artes, pero ¿qué tiene ello que ver con la cultura? El significado real de ‘cultura’, el verdadero significado de ‘cultura’, es *sensibilidad imaginativa*”. Insistió en ello a pesar del hecho de que “cultura” nunca antes había sido usada en exactamente ese sentido.

Resulta obvio que esta definición no fue una mera abreviatura, ni tampoco pretendía ser un análisis de un concepto común. Su propósito fue, por el contrario, redirigir los intereses de la gente. “Cultura” había tenido y seguiría teniendo un significado emotivo laudatorio. La definición urgía a la gente a dejar de usar el término laudatorio para referirse a la lectura y a las artes, y a usarlo, en cambio, para significar sensibilidad imaginativa. De esta manera se buscó emplazar las primeras cualidades bajo una luz más pobre y las últimas bajo una luz más brillante, redirigiendo así la admiración de la gente. Cuando la gente aprende a llamar algo por un nombre rico en asociaciones placenteras, lo admiran con más facilidad; y cuando aprenden a no llamarlo por un nombre así, lo admiran con menos facilidad. La definición usó este hecho. Cambió los intereses cambiando los nombres.

La historia previa de “cultura” facilitó este cambio. Es cierto que el significado emotivo de la palabra había crecido debido a su antiguo significado conceptual; pero *éste* estaba ahora tan firmemente establecido que persistiría aunque el significado conceptual fuera alterado de algún modo. El antiguo significado conceptual fue fácilmente alterado, pues ya se había tornado vago por el uso metafórico. Por tanto, la definición pudo efectuar un cambio en el significado conceptual sin alterar el significado emotivo. Gracias de nuevo a la vaguedad, el cambio pareció un cambio “natural”, el cual escapó de la atención de los oyentes y, por tanto, no les hizo sentirse estúpidos, pues no fueron conscientes de que estaban siendo influenciados. La eficacia de la definición se debe en parte a esto y en otra parte al hecho que hizo sus resultados permanentes al enraizarlos en los hábitos lingüísticos de las personas.

La definición puede ser llamada “persuasiva”, por tanto, en un sentido bastante convencional. Como la mayoría de definiciones persuasivas, fue de hecho doblemente persuasiva. Inmediatamente disuadió a la gente de admirar indiscriminadamente un conjunto de cualidades (la amplia lectura y el conocimiento de las artes) y los indujo a admirar otro conjunto de cualidades (la sensibilidad imaginativa). El hablante deseaba conseguir estos dos fines y, con su definición, pudo perseguir ambos al mismo tiempo.

Hay centenares de palabras que, como “cultura”, tienen un vago significado conceptual y un rico significado emotivo. El significado conceptual de todas estas palabras está sujeto a constante redefinición. Las palabras son los premios que los hombres buscan otorgar a las cualidades de su propia elección.

En el siglo diecinueve, por ejemplo, los críticos a veces remarcaron que Alexander Pope “no era un poeta”. La respuesta tonta sería que “se trata solamente de una cuestión de definición”. Es, de hecho, una cuestión de definición, pero no es “solamente” eso. La palabra “poeta” era usada en un sentido extremadamente limitado. Esto, lejos de ser algo fútil, tuvo importantes consecuencias: permitió a los críticos negar a Pope un nombre laudatorio e inducir así a la gente a que lo ignoraran. Una definición persuasiva, usada tácitamente, estaba redirigiendo sus intereses. Aquellos que deseen decidir si Pope fue un poeta deben decidir si cederán a la influencia de los críticos —es decir: si llegaran a aborrecer a Pope lo suficiente como para privarlo de un título honorífico—. Esta decisión les requerirá un conocimiento de las obras de Pope y un conocimiento de sus propias mentes. Así son las importantes cuestiones que descansan detrás de la aceptación de la restrictiva definición, tácitamente propuesta, de “poeta”. No se trata, pues, de una definición “meramente arbitraria” —tampoco ninguna definición persuasiva es “meramente arbitraria”, si se entiende que esta frase implica “adecuadamente decidido con el lanzamiento de una moneda al aire”—.

Las definiciones persuasivas son a menudo reconocidas por las palabras “real” o “verdadero” usadas en un sentido metafórico. Así, el hablante en nuestro primer ejemplo nos decía qué era la cultura “real”, distinta de la cultura “aparente”. Los siguientes son ejemplos adicionales: “caridad”, en el sentido verdadero de la palabra, significa no sólo dar oro sino dar también nuestra comprensión; el amor verdadero consiste solamente en la comunión entre mentes; “valentía”, en su sentido verdadero, es la fortaleza frente a la opinión pública adversa. Cada uno de estos enunciados es una manera de redirigir los intereses, dejando el significado emotivo de las palabras intacto y uniéndolo a un nuevo significado conceptual. De manera similar, podemos hablar del verdadero significado de “deportividad”, “genialidad”, “belleza”, etcétera. O podemos hablar del verdadero significado de “egoísmo” o “hipocresía”, usando definiciones persuasivas de estos términos despectivos que censuran en vez de alabar. Es obvio que “verdadero”, en contextos así, no es usado en el sentido literal. Dado que la gente suele aceptar lo que ellos consideran verdadero, “verdadero” toma aquí la fuerza persuasiva de “ser aceptado”. En la expresión metafórica “verdadero significado” se utiliza esta fuerza persuasiva. Se induce al oyente a aceptar el nuevo significado que el hablante introduce.

Fuera de los confines de la teoría filosófica, la importancia de las definiciones persuasivas ha sido reconocida con frecuencia. En filología han recibido

énfasis en algunas ocasiones. O, mejor dicho, a pesar de que las definiciones persuasivas han recibido poca atención, mucho se dice sobre el marco general en que se encuadraría su estudio: la interacción entre los significados emotivo y conceptual en determinar el cambio lingüístico, así como su correlación con los intereses.

Leonard Bloomfield nos presenta un ejemplo que es especialmente claro: “El promotor inmobiliario ha aprendido a apelar a toda debilidad, incluida el sentimentalismo, del posible comprador; utiliza aquellas formas de hablar cuyo contenido llevará al oyente en la dirección correcta. En sus locuciones, ‘casa’ es lo incoloro y ‘hogar’ es la palabra sentimental. De este modo el vendedor llega a usar la palabra ‘hogar’ para referirse a una caja vacía que nunca ha sido habitada, y el resto de nosotros seguimos su ejemplo”.²

Habiendo señalado que “el elemento emotivo influencia enormemente el destino de algunas palabras”, Hanns Oertel señala que “amiga” llegó a tener un sentido sinónimo con “concubina”.³ Seguro que hay muchas razones para ello. “Concubina” se había convertido en una palabra un tanto profana, demasiado fuerte para los oídos delicados. Y “amiga” permitió una cómoda ambigüedad. Siempre podía atribuirse cualquier pensamiento escandaloso a aquellos que escogieron entender la palabra en su sentido menos inocente. Pero un factor persuasivo también debe haber estado involucrado. El tacto a menudo requería que la gente se refiriera a las concubinas sin expresar desprecio. La palabra “amiga”, que, a pesar de su nuevo sentido, aún conservaba parte de su antiguo significado emotivo laudatorio, fue útil para hacer que las concubinas parecieran menos infames.

Sin embargo, las definiciones persuasivas son demasiado frecuentes como para que únicamente los filólogos se hayan dado cuenta de ellas. Un relato extremadamente perspicaz, a pesar de su giro cínico, lo da Aldous Huxley en su *Eyeless in Gaza*:

“Pero si quieres ser libre, debes ser prisionero. Esa es la condición de libertad —la verdadera libertad”.

“¡Verdadera libertad!”, repitió Anthony en una paródica voz clerical. “Siempre me han gustado este tipo de argumentos. El contrario de una cosa no es su contrario; ¡Oh, Dios mío, no! Es la propia cosa, pero como *verdaderamente* es. Pregúntenle a cualquier inmovilista qué es el conservadurismo; les dirá que es el *verdadero* socialismo. Los folletos comerciales del fabricante de cervezas están llenos de artículos sobre la belleza de la verdadera moderación. En el sentido ordinario, la moderación es tan sólo un descarado rechazo a beber; pero la ver-

² *Language* (Nueva York, 1933, 442).

³ *Lectures on the Study of Language* (Nueva York, 1902, 304-5).

dadera moderación, la *verdadera* moderación es algo mucho más refinado. La verdadera moderación es tomar una botella de vino tinto en cada comida y tres whiskies dobles después de cenar.

“Qué hay en un nombre?”, Anthony siguió, “La respuesta es prácticamente todo, si es un buen nombre. Libertad es un nombre maravilloso. Es por ello que estás tan ansioso de usarlo. Piensas que, si llamas al encarcelamiento verdadera libertad, la gente se sentirá atraída a la prisión. Y lo peor de todo es que tienes toda la razón.”⁴

II

Ya hemos insinuado que el estudio de las definiciones persuasivas se encuadra en un marco mucho más general: la correlación entre terminología e intereses. Esta correlación es altamente complicada. Unas pocas observaciones nos servirán para mostrar que nuestro análisis de definiciones persuasivas se ocupa de un aspecto muy limitado de esta correlación.

Un cambio en el significado puede ser tanto la causa como el efecto de un cambio en el interés; y las definiciones persuasivas aparecen sólo cuando el cambio en el significado es la causa. Cuando es el efecto, como cuando en los años treinta nuestra creciente desaprobación hacia las condiciones en Alemania nos llevó a usar ‘fascista’ como un epíteto, no hay en esta situación ningún elemento de persuasión; aunque una vez que la palabra ha adquirido sus asociaciones peyorativas, puede usarse más tarde en la persuasión.

Nuestro tema tiene un alcance aún más limitado que esto. Nos preocupan las *definiciones* que cambian intereses. Y es importante señalar que nos preocupan sólo *algunas* de estas definiciones. Muchas definiciones que redirigen intereses no son persuasivas. Los intereses tienden a redirigirse por *cualquier* definición, siempre que ésta cambie el significado de un término o seleccione algún sentido que excluya a otros. Cuando un científico introduce un término técnico, no importa si lo hace de manera despegada, está indicando su interés en lo que ese término nombra —es decir: que le parece importante hablar de ello o predecir su incidencia— y con frecuencia lleva a sus lectores a tener un interés similar. Sería bastante confuso llamar a estas definiciones “persuasivas”. ¿Cómo pueden, pues, distinguirse estas definiciones de las definiciones persuasivas?

La distinción depende de si el término definido tiene un fuerte significado emotivo y de si el hablante usa la palabra con carga emotiva con propósitos dinámicos —es decir: con la *intención* principal de cambiar los intereses de la

⁴ *Eyegles in Gaza* (Nueva York, 1936, 90).

gente—. Los hombres a veces dicen “No me importa qué palabra uses, siempre que hagas mi distinción”; y también “Si no estás interesado en mi distinción, está bien; dirigiré mis comentarios al limitado grupo de personas que están interesados en mi distinción”. Las definiciones dadas de una forma así no son persuasivas, pues aunque indican los intereses del hablante, y puede suceder que influyan los intereses del oyente, no hacen uso del significado emotivo en un esfuerzo deliberado para alterar los intereses.

Tal distinción es demasiado restrictiva, sin embargo, y debe ser revisada ligeramente. Cuando se da una definición con el propósito principal de distinguir o clasificar, o cuando se usa sólo para guiar aquellos intereses que (como la *curiosidad*) están involucrados en hacer la clasificación comprensible, y cuando no se sugiere que ésta es *la única* forma legítima de clasificación, entonces la definición no es persuasiva. (Esto no quiere decir que nunca se usen definiciones persuasivas en textos científicos, tampoco que las definiciones no-persuasivas estén basadas en un fundamento incuestionable, ni que las definiciones persuasivas sean menos respetables que las demás).

Debemos pasar ahora a otro punto. Las definiciones persuasivas redirigen intereses cambiando tan sólo el significado conceptual de un término con carga emotiva, permitiendo así que el significado emotivo se mantenga más o menos constante. Está claro que el cambio opuesto es igual de importante y prevalente: el significado emotivo puede alterarse y el significado conceptual mantenerse constante. Este último mecanismo no es menos persuasivo. De hecho, con frecuencia puede obtenerse la misma fuerza persuasiva mediante uno u otro cambio lingüístico. Así, en nuestro primer ejemplo de “cultura”, el hablante usó una definición persuasiva. También podría haber insistido en enunciados del tipo: “Cultura es sólo el oro de los tontos; el verdadero metal valioso es la sensibilidad imaginativa”. Este mecanismo hubiera permitido que “cultura” mantuviera su antiguo significado conceptual pero hubiera tendido a hacer su significado emotivo peyorativo, y hubiera incrementado el significado emotivo laudatorio de “sensibilidad imaginativa”. Esto hubiera servido al mismo propósito que la definición persuasiva. Las cualidades comúnmente referidas por “cultura” se colocarían del mismo modo bajo una luz pobre y la sensibilidad imaginativa se colocaría bajo una luz brillante; pero esto sucedería mediante un cambio en el significado emotivo y no en el significado conceptual.

Debemos excluir ejemplos de este último tipo de nuestra explicación de definiciones persuasivas. Aún siendo persuasivas, el efecto no se obtiene mediante definición, sino mediante los gestos y el tono de voz, o mediante mecanismos retóricos como símiles y metáforas. Es conveniente restringir la palabra “definición” a los casos en que el significado conceptual sea el único

que es determinado o, al menos, a aquellos casos en que este aspecto predomine. No debemos olvidar, sin embargo, que muchos enunciados que cambian principalmente el significado emotivo de las palabras pueden, en un sentido amplio, llamarse “definiciones”; y que, no en menor medida que las definiciones persuasivas en nuestro sentido estricto, pueden confundirse fácilmente con enunciados que no son persuasivos. (Por ejemplo, “‘conciencia’ significa la voz del destino”).

Los comentarios de las últimas páginas pueden resumirse así: las definiciones persuasivas, lejos de explicar toda la interrelación entre terminología e intereses, versan sólo sobre aquellos casos en que un cambio en la terminología *causa* un cambio en el interés, donde el significado emotivo y el uso dinámico están involucrados, y donde el cambio terminológico se encuentra sólo en el significado conceptual.

Otro comentario merece atención. Obviamente, la redirección de los intereses de la gente depende de mucho más que el significado emotivo. Depende también del uso dinámico: de la fuerza del hablante, de sus gestos, de su tono de voz, de la cadencia de sus frases, de las figuras retóricas que emplea, etcétera. Está también condicionada por el temperamento de los oyentes, su respeto hacia el hablante, su susceptibilidad a la sugestión, sus prejuicios e ideales latentes —y, de hecho, por sus creencias fácticas, pues un cambio repentino en las creencias del hombre prepara el cambio (aunque a menudo mediante un “lapsus”) para la redirección de sus intereses—. La persuasión es raramente efectiva si los oyentes no están ya predispuestos a cambiar sus intereses. Una definición persuasiva puede ser entonces importante, como un impulso final al cambio y un mecanismo mnemotécnico, enraizado en el lenguaje, que mantiene el cambio permanente. Al insistir en las definiciones, y por tanto en la función del significado emotivo, hemos enfatizado tan sólo uno de los aspectos de las situaciones persuasivas. Hay, sin embargo, excelentes razones para este énfasis. El significado emotivo es un elemento bastante estable del conjunto muy variado de factores de los que depende la persuasión eficaz y, aunque es un factor parcial, es a menudo esencial. Además, cuando un hombre redefine un término con carga emotiva, se esfuerza con *mucha* frecuencia por persuadir, y se preocupa de que se cumplan los otros factores que son necesarios para una persuasión efectiva. El significado emotivo es un *signo* claro de persuasión —permite que ésta sea detectada—. Esto es importante en el caso de las definiciones, donde la persuasión, por legítima y vital que sea en sí misma, puede adquirir muy fácilmente un atractivo espurio al enmascarse bajo la apariencia de un análisis lógico.

III

Habiendo explicado qué son las definiciones persuasivas, veamos ahora la importancia que tienen en filosofía.

Es fácil empezar considerando definiciones filosóficas de la propia palabra “filosofía”. Ramsey la define como un sistema de definiciones. Van der Leeuw como un intento de penetrar las apariencias. La divergencia entre ambos no es un accidente terminológico. “Filosofía” es un término digno, y cada hombre lo reserva para la investigación que más desea dignificar.

Consideremos la palabra “realidad”. Los filósofos no suelen buscar la realidad sino la Realidad —o, más bien, la verdadera Realidad. Pero “verdadera Realidad”, del mismo modo que “verdadera cultura”, fácilmente se define de muchas maneras diferentes, con muchos efectos persuasivos diferentes—. ¿Eran las sombras de la caverna de Platón sombras “reales”? ¿Habían sombras “reales” de caballos y hombres, distintas de las sombras imaginarias de centauros? No sirve expresarlo así. “Real” es un término demasiado impresionante para usarse en la descripción de sombras cambiantes; así que debe dársele un sentido restrictivo que lo haga predicable sólo de modelos eternos. (Cuando “Realidad” es usada por los místicos, los efectos de una definición persuasiva tácita se tornan aún más obvios).

¿Por qué Spinoza, tan ansioso de liberar al pensamiento de antropomorfismo, tentó a sus lectores al antropomorfismo usando la palabra “Dios”? ¿Por qué no habló siempre de “La Sustancia Única”? Pueden señalarse, por supuesto, las fuerzas políticas y sociales de la época, que hacían obligada una apariencia de ortodoxia. Pero ciertamente esto no es todo. La palabra “Dios” despierta, como por arte de magia, el más profundo de los sentimientos. Al dar a la palabra un nuevo significado conceptual, Spinoza pudo alejar su fuerza emocional de las viejas ficciones antropomórficas y centrarla en la Sustancia, que tan seriamente pensó que sería un objeto más gratificante para nuestra admiración y humildad. Si hubiera dicho “no hay Dios, sólo la Sustancia y sus Modos”, entonces hubiera dicho lo que él creía, dado que “Dios” sería usado aquí en su sentido popular. Pero esto hubiera sido una pobre economía de las emociones. Hubiera eliminado el objeto de admiración y humildad de los hombres sin ofrecer ningún sustituto; y estos sentimientos habrían muerto, dando lugar a un gran empobrecimiento de la vida emocional. Era necesaria la definición persuasiva de la palabra para preservar la vitalidad emotiva. El cambio de significado de “Dios” fue demasiado abrupto, sin embargo, como para no darse cuenta de ello. Tardó tiempo para que Spinoza “el ateo” diera lugar a Spinoza “el hombre intoxicado de Dios” —pues los defensores de la ortodoxia no tardaron en ver que su Dios era Dios sólo en el significado emotivo—.

Estos comentarios no deben malinterpretarse como cinismos. Identificar la persuasión no es necesariamente condenarla, tampoco debe identificarse toda persuasión con la persuasión de un orador populista. Sin embargo, es obligado distinguir entre persuasión y demostración racional.

Pasemos ahora a un tema más reciente. El Positivismo alcanzó su gran atractivo antes del “principio de tolerancia” de Carnap, y lo consiguió en gran parte gracias a su enunciado de que “la metafísica carece de significado”. ¿Pero no es este enunciado sorprendentemente parecido al de los críticos del siglo diecinueve que decían que Pope “no era poeta”? Los positivistas estaban exponiendo una verdad incuestionable de acuerdo a su sentido de “significado”, del mismo modo que los críticos del siglo diecinueve lo hacían de acuerdo a su sentido de “poeta”. La verdad de estos enunciados, sin embargo, carece completamente de importancia. La controversia depende de las palabras emotivas que se usan. ¿Debemos definir “significado” de una manera tan restrictiva para que así sólo la ciencia reciba este título laudatorio y la metafísica reciba el título peyorativo de “sin-sentido” (“*non-sense*”? ¿Debe nuestra terminología mostrar la ciencia bajo una luz más brillante y la metafísica bajo una luz más pobre? En pocas palabras, ¿debemos aceptar esta definición *persuasiva* de “significado”? Ésta es la pregunta, aunque está bien oculta por la máxima que las definiciones son “meramente arbitrarias”.

Pero esta conclusión merece una prudente aclaración. Volviendo a la analogía anterior, debemos recordar que los críticos del siglo diecinueve no condenaban a Pope simplemente con grandilocuencia. También hacían una distinción. Su restrictivo sentido de “poeta” tenía la función de enfatizar, en la atención del lector, ciertas características que eran comunes a toda poesía pero que no estaban presentes en la poesía de Pope. Quizás querían decir algo así: “Durante largo tiempo hemos estado ciegos a las diferencias fundamentales entre la obra de Pope y la de Shakespeare o Milton. Es sólo debido a esta ceguera que nos ha parecido bien dar a Pope un título laudatorio. Hagamos notar la diferencia, pues, y privémoslo de ese título”. El argumento de los positivistas puede fácilmente recibir la misma interpretación. Quizás querían decir: “Durante largo tiempo hemos estado ciegos a las diferencias fundamentales entre el uso de los enunciados en ciencia y su uso en metafísica. Es sólo debido a esta ceguera que nos ha parecido bien dignificar la metafísica con títulos como ‘significativa’ (*meaningful*). Definamos ‘significado’, pues, de una forma que claramente enfatice estas diferencias fundamentales y prive a la metafísica de su título”. Formulada así, la tesis positivista no sólo tiene calor sino también luz, y no debe ser despreciada. No obstante, quizás haya todavía demasiado calor para la cantidad de luz. Subrayar las formas en que la metafísica se ha confundido con la ciencia no es algo que carezca de utilidad; y en la medida en

que los positivistas se han dedicado a esta tarea, su “conquista de la metafísica” no ha dependido de la exhortación. Pero, ¿es cierto que sus distinciones nos han llevado más allá de la *mitad del camino* a recorrer para rechazar totalmente la metafísica? ¿Vamos a recorrer el resto del camino apoyándonos en la palabra “sin-sentido” (“*non-sense*”), definida de tal modo que otorga un significado emotivo inaceptable a la metafísica aún sin ser realmente un predicado de ésta?

Esta misma pregunta surge incluso cuando a la metafísica se le niega sólo el significado “cognitivo”. “Cognitivo” es usado aquí para significar “empíricamente verificable o analítico”, y tiene un valor exclusivamente laudatorio. La crítica positivista se reduce, pues, a esto: “Los enunciados metafísicos no son ni empíricamente verificables ni analíticos; por tanto, no son respetables”. Si los metafísicos responden “nuestros enunciados, aunque no son ni empíricamente verificables ni analíticos, son aún así respetables”, es difícil que se les pueda apartar de esta postura si no es por mera exhortación.

Los impulsos metafísicos son demasiado vigorosos para un tratamiento hortatorio; son inhibidos pero no eliminados. Si en nuestra estima la metafísica ha de ceder por completo su lugar a la ciencia, esto sólo puede venir de un examen más detallado, tanto de la metafísica como de la ciencia. Las investigaciones sobre verificación y sintaxis son un buen punto de partida, pero no son el único objeto de estudio. Sería bueno considerar como se utilizan en las ciencias aquellas palabras que sugieren imágenes gráficas y metáforas; y contrastar su función allí con su función en la metafísica; o examinar las necesidades psicológicas y las confusiones específicas que llevan a la gente a pensar que la metafísica es necesaria. Este tipo de investigaciones dirigirían nuestras actitudes hacia la metafísica de una manera más permanente e iluminadora; darían forma a nuestras actitudes al clarificar y engrandar nuestras creencias. Si una actitud adversa hacia la metafísica fuera construida de este modo, la palabra “sin-sentido” (“*non-sense*”), definida de manera persuasiva, sería útil para ayudar a cristalizar la actitud. Un programa así parece más prometedor que el de los metafísicos. Es una pena, pues, ocultar su verdadera complejidad por usar prematuramente una definición persuasiva.

IV

Pasemos ahora a la ética, con especial atención a la palabra “justicia”, tal y como es definida en la *República* de Platón.

Como recordarán, el primer libro de la *República* trata en gran parte de la discusión entre Sócrates y Trasímaco. Sócrates es el vencedor y, sin embargo, no queda satisfecho. “He ido de un tema a otro”, dice Sócrates, “sin haber des-

cubierto lo que buscaba en primer lugar, la naturaleza de la justicia. Dejé esa investigación y me puse a pensar si la justicia es virtud y sabiduría, o maldad y locura”.

¿Fue este argumento sobre la “virtud o el mal” de la justicia realmente una digresión injustificada? A la luz de nuestra discusión anterior, no podemos estar de acuerdo en que lo fuera. El argumento tenía la importante función de determinar si “justicia” debía o no retener su significado emotivo laudatorio, y esto fue esencial para el posterior desarrollo del diálogo. Cuando un hombre está a punto de dar una definición persuasiva (y en seguida veremos que Sócrates estaba a punto de hacerlo) debe asegurarse de que el significado emotivo del término definido está bien establecido. De lo contrario, una definición que pretendía iluminar un significado conceptual bajo un título laudatorio terminará oscureciéndolo bajo uno peyorativo. La palabra “justicia”, que es un tanto demasiada seria para resultar totalmente agradable, corre el peligro de tornarse peyorativa, especialmente cuando hombres como Trasímaco (con una técnica persuasiva como la mencionada anteriormente en la página 111) usan sus habilidades retóricas para *hacer* que la palabra sea peyorativa. Sócrates debía alabar la justicia, pues, antes de definir “justicia”.

La pregunta por el significado de “justicia” reaparece en el libro cuarto. Los dos libros intermedios han redirigido nuestros intereses gracias a una conmovedora descripción del Estado ideal. Estos nuevos intereses deben hacerse permanentes. Esto puede hacerse dignificando los aspectos más distintivos del Estado bajo títulos laudatorios. De los cuatro términos laudatorios que Sócrates menciona, “sabiduría”, “coraje”, “templanza” y “justicia”, los tres primeros sirven con facilidad a este propósito sin tener que experimentar un gran cambio en su significado conceptual. El término restante debe reservarse para cualquier otra cosa que necesite dignidad. Y así se encuentra la definición de “justicia”. “La justicia del Estado consiste en cada una de sus tres clases haciendo el trabajo de su propia clase”.

Difícilmente se puede negar el carácter persuasivo de esta definición —es decir: el hecho de que forma parte de un enérgico llamamiento a un nuevo sistema de clases, una inspiradora y bella especie de propaganda aristocrática—. Los significados corrientes de “justicia” deben dar lugar al “verdadero” significado, el significado que necesita la dignidad de un nombre laudatorio.

Esta explicación sorprendería a Platón como algo realmente extraño. Sin embargo, su desacuerdo es mucho menos fundamental de lo que pueda parecer en un primer momento. Sigamos su propia explicación, enfatizando aquellos puntos que guarden analogía con nuestra explicación.

Platón habría estado de acuerdo en que el significado corriente de “justicia” es tan sólo el punto de partida. No debemos modelar nuestra definición

en base a la concepción corriente de justicia sino en base a la justicia en sí misma —es decir: en base a la Idea eterna de justicia, que hemos contemplado en una vida previa al nacimiento y que ahora sólo podemos conocer mediante una esforzada *reminiscencia*—. Una definición basada en el uso corriente solamente revelaría el recuerdo imperfecto de la Idea, alcanzado por los hombres que están atrapados en el mundo de la opinión.

Este punto de acuerdo parece de poca importancia y eclipsado por la teoría de la reminiscencia. Pero observémoslo más de cerca. ¿Cómo decidió Platón si su reminiscencia era correcta? ¿La consideró correcta cuando alcanzó una concepción que satisfizo sus más íntimas y profundas aspiraciones? ¿El método dialéctico sirvió solamente para aclarar su mente para que así sus aspiraciones pudieran dirigirse hacia algo articulado? Es difícil pensar en otra respuesta. Platón aspiraba a las Ideas; pero esto no era consecuencia de algún milagroso poder de atracción que las Ideas poseyeran. Era un asunto de necesidad analítica. Todo aquello que no fuera objeto de sus aspiraciones no era llamado Idea. Si esto es así, entonces, una vez más, nuestra explicación es cercana a la suya. Si hubiera estado construyendo una definición persuasiva de manera consciente, también hubiera seleccionado, como el significado conceptual de “justicia”, el objeto de estas mismas aspiraciones. Ninguna otra cosa hubiera garantizado el nombre laudatorio. Hemos retenido los factores que llevaron a Platón a hacer su definición sin mantener el reino poético de las Ideas, cuya función, de hecho, era sólo adornar su procedimiento, y no la de alterar su resultado.

Si la obra de Platón hubiera sido menos utópica, más satírica, Platón no habría tenido reminiscencias de un reino de las ideas sino de dos. El primer reino habría sido el hogar de los dioses, tal y como es descrito en el *Fedro*, y el segundo hubiera sido el hogar del “autor del mal”, quien hace una inesperada aparición en el décimo libro de las *Leyes*. Así como las aspiraciones habrían sido el criterio para determinar el correcto recuerdo del primer reino, las aversiones habrían sido el criterio para determinar el correcto recuerdo del segundo. La teoría de la definición habría estado entonces menos limitada a los términos laudatorios. El recuerdo podría funcionar también para los términos peyorativos. Pero sería de vital importancia en el momento de definir los términos peyorativos limitar el recuerdo al segundo reino. Los más serios errores filosóficos tendrían su origen en el fracaso de recordar el reino “correcto”, donde la corrección del reino dependería del significado emotivo del término definido.

Debemos volver, sin embargo, a la definición de “justicia”. La definición de Platón era persuasiva, pero esto está lejos de ser algo excepcional. Definiciones posteriores de “justicia” son, con escasas excepciones, igualmente persuasivas. Ejercen otro tipo de influencia, por supuesto. No todos los filósofos son aristócratas. Pero ejercen una influencia.

Consideremos la definición de Bentham. “Justicia’, en el único sentido en que tiene significado[!], es un personaje imaginario, que nace por la conveniencia del discurso, cuyos dictados son los dictados de la utilidad aplicados a ciertos casos particulares”.⁵* Dicho más fácilmente, “esto es una ley justa” es una manera hipostática de decir “esta ley contribuye a la mayor felicidad del mayor número”. Dado que muchos de nosotros estamos dispuestos a dirigirnos hacia esa dirección, esta definición puede no parecernos, de forma inmediata, como una definición persuasiva. Aún así, su énfasis en solamente los números, su énfasis en contar la felicidad del hombre pobre de igual forma que la del hombre rico, marca un claro llamamiento a una mayor democracia. La definición propaga los ideales de un gran liberal.

Puede decirse que la expresión salario “justo” significa el salario que anticipa aquél que el obrero recibiría al final del proceso, mediante el concurso de las leyes de la oferta y la demanda, tan sólo si hubiera un mercado perfecto en el sentido económico. Esta definición oculta bastante bien su persuasión, haciendo parecer que es resultado de una economía puramente científica. Pero es un llamamiento, aunque ligeramente comprometido, al funcionamiento, no de las leyes económicas, sino de las leyes económicas “naturales” —esto es: un llamamiento al funcionamiento de las leyes económicas tal y como *podrían* formularse *si* los aspectos puramente competitivos de la industria, el “sálvese quien pueda”, fueran garantizados—. Por ello encontraréis que esta definición resulta más agradable a aquellos que prosperan bajo las presentes condiciones industriales que a aquellos que no prosperan.

“Justicia” puede definirse de muchísimas maneras, siempre sin escandalizar a los lexicógrafos. ¿Ojo por ojo y diente por diente? ¿Respetar los contratos, solamente? ¿La voluntad del rey? ¿La distribución de la riqueza social de acuerdo con la cantidad de *trabajo* que cada hombre lleva a cabo? Tenemos una amplia variedad de significados y libertad, dentro de unos amplios límites convencionales, de inventar nuevas definiciones. Qué significado escogemos, sin embargo, no es un asunto trivial, pues dignificaremos ese significado con un título laudatorio. Escoger un significado es tomar partido en una lucha social.

Es curioso descubrir que los teóricos han estado todos ellos desconcertados ante la incertidumbre de la ética, y han entrevisto, incluso en momentos de calma filosófica, los elementos de persuasión que están involucrados. Han procurado evitar esta situación definiendo sus términos, con la esperanza de dar mayor rigor y racionalidad a sus investigaciones. Sin embargo, y aunque resulte irónico, estas definiciones involucraron la misma persuasión, y de un modo que la enmascaraba y la confundía, haciéndolas parecer análisis puramente intelectuales.

⁵ *Principles of Morals and Legislation* (1789, cap. 10, secc. 40, n. 2).

* *Nota del traductor*. La exclamación entre corchetes es de Stevenson.

V

Los ejemplos que hemos considerado, ya sean de la metafísica, la teología, la epistemología o la ética, muestran que las definiciones persuasivas están lejos de ser algo raro en filosofía y que el fracaso a reconocer su carácter persuasivo ha sido el responsable de mucha confusión. ¿Pero cuál es, en esencia, la naturaleza de esta confusión? En gran medida es ésta: la incapacidad de reconocer la persuasión ha fomentado un malentendido acerca del *tipo de desacuerdo* que motiva muchas discusiones, y en consecuencia ha llevado a la gente a dar soporte a sus enunciados con un *método* demasiado simple, o a buscar un método definitivo de probar sus enunciados cuando ninguno era posible.

Hasta este momento, estas confusiones metodológicas han sido evidentes sólo indirectamente; ahora debemos tratarlas de un modo más explícito. Empecemos señalando la complejidad *real* en la metodología que las definiciones persuasivas introducen, pues así la medida en que esta complejidad ha sido pasada por alto resultará entonces obvia sin añadir más explicaciones. Será conveniente limitar nuestra atención al ejemplo de “justicia” —aunque debe recordarse, por supuesto, que las mismas consideraciones aplican a cualquier caso que involucre un término que esté sujeto a una definición persuasiva—.

El resumen de la metodología será similar al dado en otro ensayo previo.* Sin embargo, el patrón de análisis que allí ejemplifica “bueno” es ligeramente diferente al que ejemplifica aquí “justicia”. Reaparecen las mismas consideraciones metodológicas, pero debemos reconocerlas bajo una nueva apariencia y en medio de complicaciones adicionales.

Dos hombres no están de acuerdo sobre si cierta ley es justa. Examinemos las diferentes formas que su discusión puede tomar.

(1) Supongamos que ambos usan “justo” con el mismo significado conceptual —a saber: *que tiene las consecuencias A y B*—. La discusión puede solucionarse usando el método empírico. Los participantes en la discusión sólo tienen que comprobar si la ley tiene estas consecuencias.

Sin embargo, éste es un caso sencillo que se da raramente. Hemos visto que “justicia” está constantemente sujeta a la definición persuasiva, con el resultado de que diferentes personas llegan a usar dicha palabra en diferentes sentidos.

(2) Supongamos que el primer hombre usa “justo” para referirse a A y a B, y el segundo hombre lo usa para referirse a B y a C. Supongamos además que B es el único punto en desacuerdo. En este caso, ambos participantes seguirán sin darse cuenta de sus diferencias en terminología y de nuevo encontrarán adecuado el método empírico. El resultado de la discusión dependerá de si se descubre o no que la ley lleva a B.

* *Nota del traductor.* Stevenson se refiere aquí a su “The Emotive Meaning of Ethical Terms”.

(3) Supongamos ahora exactamente lo mismo, con excepción de que C, y no B, es el único punto en desacuerdo. Probablemente ambos participantes se darán cuenta de su diferencia en la terminología. Aún así, la discusión puede seguir y *en algunos casos* puede ser zanjada empíricamente. Si el segundo participante, que usa “justo” para referir a B y a C, es quien niega la justicia de la ley, su oponente puede refutarle mostrándole de forma empírica que la ley lleva a C. (Hemos concedido, por hipótesis, que ambos están de acuerdo en que se da B). “Has sido refutado”, dirá el primero, “incluso de acuerdo a tu errónea concepción de justicia”.

Este ejemplo plantea un punto que requiere de una atención especial. El primero de los participantes en la discusión no refirió a C en su enunciado inicial, y el segundo de los participantes negó la justicia de la ley teniendo sólo en cuenta C. Por tanto, el enunciado inicial del primer hombre no fue contradicho en ningún momento por su oponente. Aún así, el primer hombre sentirá, incluso después de darse cuenta de la discrepancia en la terminología, que ha sido cuestionado desde el inicio. Sentirá la necesidad de refutar el enunciado de su oponente, como si ello fuera necesario para respaldar el suyo propio. ¿Por qué sucede esto?

Si la pregunta parece desconcertante es tan sólo porqué hemos tomado en cuenta exclusivamente el significado conceptual. Hemos supuesto tácitamente que los participantes en la discusión eran como científicos, motivados por una curiosidad imparcial. Si tomamos nuestro ejemplo como un caso típico en la mayoría de situaciones reales, entonces este supuesto carece completamente de justificación. El uso de “justo” e “injusto” indican claramente que uno de los participantes estaba *a favor* de la ley y el otro *en contra* de ella. Discutían por este motivo, y no porqué tuvieran una mentalidad científica. Estaban *en desacuerdo en interés*. Cada uno de ellos tenía un tipo de interés diferente en la ley, y ninguno de ellos estaba satisfecho con dejar que el interés del otro siguiera igual.*⁶ Este tipo de desacuerdo es más evidente desde el significado emotivo que desde el significado conceptual. El hecho, pues, que el significado conceptual del primero de los participantes no fuera contradicho no impidió que sintiera que su enunciado estaba siendo cuestionado. No sólo quería que su oponente reconociera ciertas consecuencias de la ley, sino que quería también que la alabara; y su oponente no alabaría la ley si la llamaba “injusta”, fuera cual fuera el significado conceptual que le asignara.

⁶ Véase los ensayos “The Nature of Ethical Disagreement” y “The Emotive Meaning of Ethical Terms”.

* *Nota del traductor*. Como señalé anteriormente, “The Emotive Meaning of Ethical Terms” fue publicado originalmente en el número 46 de la revista *Mind* (1937). “The Nature of Ethical Disagreement” [“La naturaleza del desacuerdo ético”] fue escrito en 1941 y publicado por vez primera en el número 8-9 de la revista *Sigma* (1948). Ambos ensayos están incluidos en la compilación *Facts and Values: Studies in Ethical Analysis* (1963).

El desacuerdo en interés se ve más fácilmente en casos como (3), aunque una breve reflexión muestra que está igualmente presente en los casos (1) y (2). El uso del término laudatorio “justo” en los casos anteriores muestra que los participantes en la discusión también estaban preocupados acerca de si la ley debía ser alabada o no. Las consecuencias A, B y C estaban involucradas, por supuesto, pero no lo estaban más que en el tercer ejemplo, y eran relevantes por la misma razón —es decir: relevantes porque el desacuerdo en interés, que es lo que motivó la discusión, tenía su origen en un desacuerdo en la creencia—. En otras palabras, los participantes tendrían el mismo tipo de interés en la ley tan sólo con resolver sus creencias opuestas sobre las consecuencias de dicha ley. En los dos primeros casos, estas creencias opuestas versaban sobre consecuencias que *ambos* participantes referían conceptualmente con la palabra “justo”. Este es el principal punto de diferencia entre los casos y es una diferencia que carece de importancia. El desacuerdo era de tal tipo que hubiera terminado sólo cuando ambos participantes tuvieran el mismo interés en la ley. Las creencias eran relevantes sólo en la medida en que redirigían los intereses. Qué creencias hacían eso, y si estaban expresadas en los enunciados iniciales de ambos participantes, determinaba tan sólo la complejidad de la discusión, pero no su naturaleza.

Estos comentarios nos preparan para otro caso:

(4) Supongamos, como en los casos anteriores, que el primer hombre usa “justo” para referir a A y a B, y que el segundo hombre (que niega la justicia de la ley) usa “justo” para referir a B y a C. Supongamos, además, que ambos han probado de forma concluyente que la ley lleva a A y a B, y que no lleva a C. En términos conceptuales, no han localizado ningún punto de desacuerdo, ni tampoco hay la posibilidad, como en el caso (3), de que un hombre refute al otro “incluso de acuerdo a su errónea concepción de justicia”. No obstante, aún pueden discutir acerca de la justicia de la ley. La fuerza laudatoria de “justo” y la fuerza peyorativa de “injusto” siguen siendo indicativos de un desacuerdo en interés.

Por lo que refiere a la metodología, este caso tiene una importancia particular. Representa un desacuerdo que *el método empírico puede ser completamente incapaz de resolver*.

Esto resultará claro si volvemos a considerar, aunque con ello nos repitamos parcialmente, por qué el método empírico *fue* decisivo en el primero de los casos. En cada uno de los otros casos anteriores, el juicio inicial de uno de los participantes era falso. Esto era así o bien por la ley de contradicción o bien por hipótesis explícita. Cada participante, además, tenía un interés favorable en la ley sólo en la medida en que creía que “justo”, en su sentido, era verdaderamente un predicado de la ley; pues de otro modo hubiera usado el

término laudatorio en un sentido conceptual diferente. Por estas razones los participantes tenían sólo que comprobar la verdad de sus enunciados iniciales, y esto les llevaría a tener el mismo tipo de interés en la ley. En pocas palabras: el desacuerdo en interés, que era el motivo principal de la discusión, tenía su origen en un desacuerdo en creencia —es decir: en alguna creencia que al menos uno de los participantes había expresado falsamente en su enunciado inicial—. El método empírico, al cuestionar esta creencia, podría así resolver el desacuerdo en interés.

En el caso (4), sin embargo, ambos enunciados iniciales de los dos participantes son verdaderos. Ambos participantes están dispuestos, como en los casos anteriores, a aprobar o desaprobado la ley dependiendo de si “justo” e “injusto”, en los sentidos dispares en que los usan, son verdaderamente predicables de la ley; pero una investigación empírica servirá para *dar soporte a ambos* enunciados. Por tanto, el primer hombre continuará llamando a la ley “justa”, con aprobación, y el segundo “injusta”, con desaprobación. Su desacuerdo no tiene su origen en alguna creencia que alguno de ellos esté expresando, y puede deberse solamente a los diferentes temperamentos de cada uno de los participantes. Dado que el método empírico sólo altera los intereses al alterar creencias, ¿cómo puede usarse para resolver este desacuerdo?

Es obvio que el método empírico no tiene la misma aplicación directa en (4) que en los casos anteriores. Con todo, sería una conclusión demasiado precipitada si dijéramos que no hay ningún lugar para el método empírico. Analicémoslo con más detalle.

Si el caso (4) continua en disputa, las definiciones persuasivas, que hasta ahora sólo han sido responsables de la ambigüedad de “justo”, pasarán a jugar un papel más discernible e importante. Cada uno de los participantes, con tal de influenciar el interés del otro, insistirá en su propia definición. Discutirán si la ley es justa en el sentido *verdadero* de “justo”. Hasta que no se pongan de acuerdo en el sentido de la palabra, no estarán de acuerdo en el asunto fundamental —a saber: si la ley ha de ser descrita con un nombre que merezca sus alabanzas—.

El método empírico, por inútil que pueda ser en alterar la verdad de las predicaciones conceptuales que los participantes hicieron al inicio, puede reaparecer como un medio para respaldar sus definiciones persuasivas. El segundo participante, por ejemplo, puede descubrir que C, que él refiere como “justo”, tiene otras consecuencias, F, G y H. Si tiene un interés desfavorable hacia estas consecuencias, es posible que ya no quiera definir “justo” en términos de C. Si descubre que A tiene otras consecuencias, I, J y K, hacia las que él tiene un interés favorable, puede decidir usar “justo” para referir a A. En otras palabras, puede aceptar la definición en la que su oponente ha estado insistiendo. Am-

bos participantes estarán entonces de acuerdo en que la ley es justa en un sentido de “justo” mutuamente aceptado. Este sentido será producto de su nuevo conocimiento empírico y pondrá fin a la discusión, no sólo porqué ambos crean que este sentido es verdaderamente predicable a la ley, sino porqué su aceptación mutua muestra que ya no están en desacuerdo sobre el interés, sino que ambos participantes aprueban la ley.

La discusión en (4) *puede* ser resuelta, pues, de un modo empírico, pero debemos recordar que también puede no ser así. Incluso si ambos participantes conocen todas las consecuencias relevantes de la ley, uno de ellos puede seguir deseando alabarla y el otro condenarla. No llegarán a ningún sentido conceptual común de “justo”, y a pesar de que ninguno de ellos estará necesariamente diciendo algo falso sobre la ley, continuarán estando en desacuerdo acerca de su justicia. El desacuerdo será un desacuerdo en interés, no basado en ningún tipo de desacuerdo en creencia. Si de algún modo es resuelto, sólo puede ser resuelto por exhortación.

Es una verdad general que el método empírico puede resolver un desacuerdo ético, o cualquier otro tipo de desacuerdo en interés, sólo cuando éste está basado en un desacuerdo en creencia. El presente esbozo de metodología se ha tornado complicado sólo en lo que refiere a *qué* creencias están en la base del desacuerdo en interés —es decir: si hay alguna, y si es así, en qué medida es expresada en los juicios iniciales—. Estas consideraciones son esenciales para clarificar la naturaleza de la discusión, pero no tienen una importancia adicional. Esto es obvio por el hecho de que los argumentos de este tipo emergen más del significado emotivo de los juicios iniciales que de su significado conceptual. Hay otra consideración que también hace evidente este punto: en la práctica real se usa “justo” de un modo tan vago que ninguno de los participantes en la discusión está seguro de qué consecuencias están incluidas en la definición de “justo”, y que son las que le guían en el proceso psicológico de formar sus definiciones de “justo”.

El presente modelo de análisis se aplica de forma adecuada al resto de los términos éticos más específicos, así como a “bello”. El modelo de análisis ejemplificado en otro lugar por “bueno”^{7*} se aplica de forma adecuada sólo a los

⁷ “The Emotive Meaning of Ethical Terms”, sección 4. Véase también el ensayo “Retrospective Comments”, sección 9, donde se presenta una visión un tanto diferente por lo que refiere a los “términos éticos más específicos”.

* *Nota del traductor.* El ensayo “Retrospective Comments” [“Comentarios en retrospectiva”], redactado en 1963, es el último de los ensayos incluidos en la compilación *Facts and Values: Studies in Ethical Analysis*. En él Stevenson ofrece algunos comentarios y actualizaciones a sus ensayos anteriores. El fragmento al que hace referencia aquí Stevenson dice así: “Consideremos ‘él es valiente’, donde ‘valiente’ produce un juicio (*prima facie*) favorable que es más bien específico. A grandes rasgos podemos tomarlo como equivalente a ‘habitualmente planta cara al

términos éticos más genéricos. (No ofrece ningún medio que sea rápido para indicar *differentiae*). Pero cuál de estos modelos de análisis seleccionamos para un término ético es en gran parte un asunto de conveniencia técnica. Quizás podría tratarse “justo” del mismo modo que “bueno” y distinguirse de “bueno” por el tipo de interés que involucra —aunque la terminología psicológica actual no ofrece los medios para hacer esta distinción de manera precisa—. No hay dudas de que “bueno” podría tratarse del mismo modo que “justo”. Moritz Schlick comenzó a hacer esto,⁸ pero su falta de énfasis en el desacuerdo en interés, y todo lo que éste implica, vicia en gran parte su explicación. Lo mismo

peligro de un modo que es admirable’. [...] Es evidente que la fuerza evaluativa de ‘valiente’ es preservada solamente por ‘admirable’. La primera parte del análisis que he propuesto no tiene fuerza evaluativa, sino que simplemente ofrece una descripción fáctica medio-indeterminada del hombre en cuestión. [...] Así, según el análisis que he propuesto, ‘él es valiente’ es de carácter híbrido: su significado tiene a la vez un importante componente fáctico y un importante componente evaluativo. Y no cabe duda, creo, que es el componente fáctico más que el componente evaluativo el que hace que su significado sea específico y no genérico. Este ejemplo no es ajeno a mi ensayo ‘Definiciones persuasivas’, donde expliqué que una definición persuasiva puede en principio causar que cualquier juicio de valor contenga un componente fáctico. Pero este ejemplo enfatiza un punto que en el ensayo ‘Definiciones persuasivas’ no se hizo suficiente explícito. El componente fáctico de ‘él es valiente’, en la medida en que incluye una u otra manera de habitualmente plantar cara al peligro, *pertenece* simple y llanamente a la oración, y es fijado por las convenciones de nuestro lenguaje. Si intentamos eliminarlo mediante una definición persuasiva —por ejemplo, ‘es inevitable que un hombre verdaderamente valiente se paralice ante la presencia de un peligro’—, dejaremos al resto perplejos. En general, un juicio híbrido (que puede llamarse igualmente un *enunciado* híbrido, aunque podría llamarse más gustosamente un *comentario* híbrido) tiene una fuerza evaluativa que conlleva un ‘núcleo’ de descripción fáctica, siendo éste no solamente permitido por las reglas del lenguaje, sino que es de hecho requerido por ellas. El núcleo de descripción fáctica complica las razones que sustentan tales juicios, pero sólo las complica en aspectos obvios. Al considerar si un hombre es valiente, por ejemplo, debemos determinar en primer lugar si habitualmente planta cara al peligro. Esta es una pregunta claramente empírica. Si nuestra observación muestra que habitualmente no planta cara al peligro, entonces el juicio se tornará insostenible. Pero si muestra que lo hace, entonces el juicio será establecido solo parcialmente; [...] aún quedará la pregunta de si esas maneras en que habitualmente planta cara al peligro son realmente *admirables*. A partir de aquí estaremos tratando con valores; y nuestras razones, que anteriormente se habían relacionado con *el juicio* inductivamente, ahora se relacionarán con él de otro modo. Ya no son razones para creer que el hombre tiene ciertas características, sino razones a favor (o en contra) de admirarlas. [...] Por supuesto que aquí, como en cualquier otro ejemplo que involucre valores, podemos introducir definiciones persuasivas. Esto sólo quiere decir que pueden añadirse temporalmente algunas de las razones del último tipo que he mencionado al significado de ‘admirable’, y que esas razones pueden, por tanto, añadirse también al significado de ‘valiente’. Nótese, sin embargo, que pueden añadirse sólo temporalmente, y sólo para este tipo de contexto. Pueden de nuevo ‘restarse’ fácilmente por cualquiera que rechace la definición persuasiva. Es por ello que no debemos confundir el componente fáctico que introduce una definición persuasiva con el núcleo del significado fáctico que establece las reglas fijas del lenguaje.” (*Facts and Values: Studies in Ethical Analysis*. New Haven, Yale University Press, 1963, 221-3).

⁸ *Fragen der Ethik* (Viena, 1930, cap. 1).

puede decirse, aunque con más reservas, de la explicación dada por Ogden y Richards⁹ y de la explicación dada por C. D. Broad.¹⁰

Los términos éticos se usan tan vagamente que muchos modelos diferentes de análisis son relevantes para su uso convencional. Es inútil señalar uno de ellos como *el* patrón de análisis. Todo lo que se requiere es que el análisis clarifique, de un modo u otro, las características esenciales de las discusiones éticas. Éstas son el significado emotivo, el uso dinámico, el desacuerdo en interés y el papel, importante pero no definitivo, del método empírico.

⁹ *The Meaning of Meaning*, 149.

¹⁰ "Is Goodness the Name of a Simple, Non-Natural Quality?". *Proceedings of the Aristotelian Society*, n. s. 34 (1933-34, 249-68).

Notes de discussió i ressenyes

ONCINA COVES, Faustino (ed.) 2019, *Crítica de la Modernidad. Modernidad de la crítica (Una aproximación histórico-conceptual)*, València: Pre-textos. ISBN 978-84-17830-98-4, 181 pàgines.

Amb la *Modernitat* s'engega el temps de la *crítica*, que és erigida com a instància fonamental del pensament; ara bé, en el mateix moviment, la *crítica*, forçosament, prendrà com a objecte a la pròpia *Modernitat*. El diagnòstic actual d'aquesta parella conceptual és el que convoca les diferents intervencions que compila el llibre editat per Faustino Oncina Coves que ací ressenyem, *Crítica de la Modernidad. Modernidad de la crítica*, que troben la seua base en comunicacions d'un congrés homònim. La inspiració d'aquestes dues veus —“crítica” i “Modernitat—, ens assenyala Oncina aclarint l'enfocament del diagnòstic, la localitzem en el diccionari *Conceptes històrics fonamentals. Lèxic històric del llenguatge politicosocial a Alemanya* editat per Werner Conze, Otto Brunner i Reinhart Koselleck, títol fonamental de la història conceptual [*Begriffsgeschichte*]. Tanmateix, els diccionaris històrico-conceptuals canònics requereixen d'una actualització: bona part del nostre patrimoni lingüístic ja no hi queda consignat. Inscrit en tan ambiciós projecte actualitzador, aquest volum retrata la temptativa d'una història conceptual renovada d'alambinar el concepte de crítica —i el seu vincle “crític” amb la Modernitat— en diàleg amb diverses perspectives amb avalada experiència en l'esmentada problemàtica. Es tracta d'un text realment plural. El lector notarà les evidents discordances, tant de punt de partida com d'objectius, entre els diferents autors. Tanmateix, lluny de constituir un punt feble, aquest caràcter virolat del conjunt d'aportacions mostra la complexitat i les contradiccions del propi objecte.

Passem ara a presentar els capítols. Seguirem, per guiar-nos, un fil argumental que els farà aparèixer en un ordre diferent al del llibre.

Precisament del caràcter amfibològic del concepte de “crítica” parteix Oncina en el capítol inaugural amb la intenció de fer un exercici clarificador. Delimita, en primer lloc, les diverses accepcions de la “crítica” i, seguidament, configura els distints aspectes de la “crítica de la ideologia”. Aquest exercici li permet, a banda de situar la problemàtica, apuntar a la idoneïtat de Koselleck: en ell trobem, en les seues nocions o connotacions, la pluralitat reunida sota ambdós conceptes. És per això que la variant koselleckiana de la història conceptual serà la més fructífera. Oncina esclareix en primer lloc l'ús que fa

Koselleck de la crítica a *Crítica i crisi*, la seua tesi doctoral. Vinculada a la inescindible parella, la “crisi”, la crítica apareix ací concebuda com la forma que pren el pensament en la Il·lustració. La crítica també compareix en l’elaboració metodològica del lèxic *Conceptes històrics fonamentals*, en la primera de les pautes metodològiques que enumera Koselleck, la “crítica històrica”. També anomenada “crítica de fonts”, aquesta es proposa elucidar els conceptes d’un temps determinat no només semànticament, sinó també pragmàtica. Seguidament, tot i que la qüestió ja està inoculada des de l’inici, Oncina assenta les bases per a assajar el diàleg entre la Història conceptual i la crítica de la ideologia. L’èmfasi es col·loca en la koselleckiana caracterització dels conceptes com a *indicadors d’experiències* i *factors d’expectatives*, així com el seu precepte d’indagar els estrats semàntics dels conceptes vinculant-los amb la dimensió pragmàtica del seu ús. Seguint l’estela de l’últim Koselleck, conclou Oncina, la història conceptual pot exercir-se avui com a crítica ideològica de les noves formes de mixtificació —desincronització de les esferes socials i aparició en la contemporaneïtat d’allò no contemporani com a mercaderia—.

També Ernst Müller, al segon capítol, tracta d’escatir la forma en què la crítica és posada en moviment en la història conceptual. La trobem en la seua primera i més bàsica tasca, en el mètode “històrico-crític” (hi al·ludíem amb Oncina). En aquest moment purament sincrònic, la dilucidació dels significats històrics d’un concepte permet descobrir significats devaluats o reprimits, reportar anacronismes. Però també mantenim la crítica en la segona instància en què, introduint la diacronia, elevem el mètode històrico-crític a història conceptual pròpiament dita. Ací, aquesta pot servir com un correctiu al necessari gest d’abstracció del context que ella mateixa efectua en seguir el significat dels conceptes en el temps. Davant l’aparença de constància que adquireixen els conceptes pot, en suma, apropar-nos-els com accions humanes que responen a desafiaments concrets. Una història conceptual així concebuda possibilita la connexió crítica amb el passat reprimat en el sentit de W. Benjamin. A més, també pot posar de manifest els pressupòsits pràctics de la ciència a través de la reconstrucció de les preguntes i els problemes en la superació dels quals treballaren els conceptes. Ara bé, la història conceptual, per poder ser crítico-ideològica, ha de tenir en compte “allò Altre de la semàntica”. Müller explicita el distanciament de Koselleck dels seus mestres, Heidegger i Gadamer, quan assenyalava que l’anàlisi dels conceptes no ha de veure’s clausurada en la seua consideració com a llenguatge, sinó que ha d’apuntar al seu caràcter d’índexs d’una lògica social general. Partint d’aquesta precisió i valent-se del contrast amb Koselleck, Müller critica a Harmut Rosa justament per sobredimensionar la semàntica sobre els processos socials quan tracta l’ “acceleració”, o el seu parell antagònic, la “ressonància”, com a metafenòmens sociològics.

L'intent de trobar en la història conceptual la crítica (ideològica) l'apropa en les interpretacions a la tradició de la Teoria Crítica frankfurtiana; en part ho hem vist. Nogensmenys, aquest diàleg requereix, per ser fructífer, d'una explicitació de les diferències d'ambdós projectes i dels possibles elements que suposen una resistència a l'apropament. Serveix per a aquest propòsit el recordatori d'Oncina, al primer capítol, del fre que suposa la transcendentalització de categories arrelades en un moment històric determinat de la *Històrica* kose-lleckiana per a la crítica ideològica (malgrat l'autocrítica de Koselleck al respecte). També cal tenir present que la crítica a la Modernitat que efectua la història conceptual en algunes de les seues elaboracions clàssiques és, per conservadora, radicalment oposada a la dels frankfurtians (tot i coincidir en assenyalar el seu caràcter dialèctic). L'estudi que elabora Bárbara Picht en el capítol quart exposa justament una d'aquestes crítiques, la de Werner Conze. Per a ell, la "revolució mundial europea" ha pertorbat l'ordre social, sent el distanciament entre Estat i societat l'avatar principal d'aquest aldarull modern. Davant la desintegració a què l'acceleració moderna ha abocat, amb una mirada funcionalista, Conze s'interessa per com efectuar una reintegració renovada de l'ordre que continga el desbocament dels elements exclosos —vegeu, fonamentalment, la classe treballadora—. Com s'observa, es tracta d'un diagnòstic eminentment polític: la investigació històrico-conceptual és considerada com una forma d'il·luminar l'"obscuritat moderna". La tríada present en l'anàlisi de Conze, a saber, *món integrat-dissolució-reintegració renovada*, fou usada també, recorda Picht, per historiografies com la *Ostforschung* que justificaren l'expansió nacional-socialista cap a l'Est xifrant-la com una construcció d'un nou ordre. Conze guardà un "prudent silenci" respecte d'aquestes formes de trànsit de la història conceptual a la política conceptual, però no fou el cas d'altres com Otto Brunner, qui posà la crítica històrica a disposició del model *völkisch* de pensament.

Són diversos els capítols centrats en la Teoria Crítica. La història conceptual la trobem ací nodrint metodològicament el rastreig dels distints significats que la "crítica" ha anat adoptant en els diferents herois de l'Escola de Frankfurt. Vanessa Vidal reconstrueix en el cinquè capítol el concepte d'*interpretació* que Adorno elabora en la seua lectura de Kierkegaard. Aquesta s'ha de comprendre, sosté Vidal, com una particular forma d'apropiació del materialisme i del projecte de la crítica diferent i enfrontada al "materialisme interdisciplinari" de Horkheimer, identificat tradicionalment aquest últim com el genuí projecte de la Teoria Crítica. Interpretant la literalitat del llenguatge kierkegaardian, tal i com ho disposa la teoria freudiana del llenguatge, Adorno aconsegueix que aquest aparega en tant que símptoma o indicatiu no de continguts subjectius inconscients latents, com en els actes fallits freudians, sinó d'allò que en el vocabulari hegelian és l'esperit objectiu. En la literalitat dels textos de Kierkegaard

s'expressa la figura de l'esperit objectiu de l'època, però —i aquest és el punt fonamental on Adorno marca la seua distància amb Hegel— no com a realitat racional, sinó en el seu caràcter mític, absurd. Precisament perquè fa aparèixer en el seua irracionalitat a la “segona naturalesa”, la interpretació adorniana connecta amb la praxis transformadora.

Si Vanessa Vidal desmarca el projecte de teoria crítica del jove Adorno del de Horkheimer, el capítol sisè ens du, de la mà de Falko Schmieder, a un abordatge del concepte de “crítica” —així com del projecte de Teoria crítica associat— a partir de Horkheimer. Ací la història conceptual il·lumina, junt amb la història dels problemes, a la història de la teoria per tal de mostrar com el concepte de “Teoria Crítica” és l'índex d'un problema concret al qual s'enfronta el pensament de filiació marxista en els anys trenta, a saber, la crisi del vincle entre teoria i praxis. L'anàlisi discorre a partir de la comparació de dos textos: el discurs inaugural de l'Institut d'Investigació Social (1931) i el text programàtic *Teoria tradicional i Teoria Crítica* (1937). En el primer, Horkheimer dissenya un programa d'investigació social compatible, com en Marx, amb una concepció forta de ciència vinculada amb la confiança de què la teoria pot introduir racionalitat en el subjecte i el món. En contrast, aquesta confiança es dissipa en el text de 1937 (entretant ha ascendit Hitler al poder, l'Institut s'ha vist obligat a prendre el camí de l'exili i l'Est s'ha estalinitzat). El concepte de “Teoria Crítica” del Horkheimer de 1937 implica una “rehabilitació de la filosofia” que, com deia Adorno, es manté en vida perquè deixà passar l'instant de la seua realització. Aquest problema que hi ha a la base del concepte de Teoria Crítica és oblidat en la seua reaparició en els anys seixanta.

També rastrejant el concepte de crítica, però en un sentit diferent, trobem la detinguda exploració que José Manuel Romero Cuevas efectua en el novè capítol d'allò que d'específic té la crítica de la Modernitat en els principals representants de l'Escola de Frankfurt. Romero seguirà el rastre de la *crítica immanent*, no només en Adorno i Horkheimer, sinó també en els seus “successors”. La ja tradicional lectura iniciada per Habermas reprova a Adorno i a Horkheimer haver despullat a la crítica de fonaments racionals com a conseqüència d'identificar a la raó realitzada en la nostra civilització com a raó instrumental; la resolució d'aquesta aporia és un mèrit que s'apuntarien la teoria habermasiana de l'acció comunicativa i la teoria del reconeixement d'Axel Honneth. En litigi amb aquesta reconstrucció, Romero pretén assenyalar la potència de la proposta d'Adorno i Horkheimer, alhora que critica els plantejaments dels seus crítics qüestionant el suposat guany que suposen respecte dels mestres. A llum del diagnòstic, apareixen Adorno i Horkheimer com a defensors d'una crítica immanent vinculada amb la crítica al capitalisme que serà inexistent en Habermas i que haurem d'esperar a l'últim Honneth i a Jaeggi perquè reaparega.

Un exercici particular és el que elabora Giorgia Cecchinato al vuitè capítol. Amb ànim de donar concreció a aquestes reflexions teòriques, posa la concepció de W. Benjamin de la crítica, així com les seues lectures de Kafka, a disposició de l'abordatge d'un fet actual: l'intent d'ocultació de la memòria de la dictadura al Brasil. Benjamin vincula la interpretació amb el portar el passat al present de forma que l'últim pugui obrir-se en un sentit transformador. L'adopció d'aquest enfocament permet a Cecchinato defensar la urgència de revisar la Llei d'amnistia brasilera de 1979 i de restituir les restes mortals dels desapareguts com una forma d'elaboració del passat, de “donar sentit a la ferida” —signe de l'oblit i, al seu torn, de la impossibilitat d'oblidar—. Aquest no és un gest de mer record del passat, sinó de lluita per la transformació del present.

Altres abordatges crítics de la Modernitat que la història conceptual fa bé a tenir en compte per la seua transcendència són els de Hegel i Heidegger. Al capítol tercer, Giovanna Pinna elabora una reflexió sobre la filosofia hegeliana de l'art que treu a la llum les seues potencialitats com a crítica de la Modernitat. Pinna mostra com en ella Hegel posa en dialèctica relació els aspectes de la “Modernitat com a crítica” i de la “crítica de la Modernitat”. Pel que fa al primer, Hegel identifica l'especificitat d'allò modern en la ruptura respecte de la continuïtat històrica, en la consideració del passat com un “altre” permesa per la posició reflexiva adoptada. Així, en les novel·les de cavalleries —fase intermèdia entre allò antic i allò modern— encara trobem la “clàssica” adhesió a l'univers espiritual que representa l'obra. En canvi, en *El Quixot*, reelaboració irònica de la bella cavalleria, la distància entre la matèria tractada i l'elaboració artística és una marca d'allò plenament modern. L'aspecte de la “crítica de la Modernitat” apareix associat al concepte de *prosa del món*, amb el qual Hegel fa referència al desencantament de la quotidianitat —vinculat amb la racionalització moderna— que empenya l'art a l'ús de la comicitat i la ironia. Cervantes és modern perquè ja no pot representar l'heroi, que antuvi carregava als muscles amb allò universal i era la pauta moral, sinó de forma irònica —sent la ironia una reacció a l'espai limitat d'acció del ciutadà modern—. Des d'aquesta perspectiva trobem un Hegel que ens avança la crítica weberiana a la Modernitat com a pèrdua de sentit.

Un dels caràcters fonamentals de la Modernitat és que té una concepció forta del subjecte. Des de finals del segle XIX s'han efectuat potents crítiques al subjecte que són, no cal dir-ho, crítiques a la Modernitat. Juan de Dios Bares Partal efectua al setè capítol una anàlisi d'una d'aquestes: la de Heidegger. Ho fa a partir de, per una banda, el text conegut com *l'Informe Nartop* (1922) i, per l'altra, *L'èsser i el temps* (1927). Heidegger s'embarca en allò que ell denomina *Destruktion* de la tradició, una revisió de com s'ha comprès a l'èsser des

del temps, o de l'ésser i el temps, que el duu a remuntar-se a les fonts. Una de les fonts fonamentals és Aristòtil. En aquesta *Destruktion* ambdós textos corren paral·lels. *L'Informe Naptop* valora en Aristòtil una concepció de l'ésser vinculada amb la pràctica, que remet a “allò que està a la mà”. En canvi, *L'ésser i el temps* menysté la tradició i assenyalava els límits de l'ontologia d'un Aristòtil que estaria impregnat de la metafísica de la presència. No sense passar reticències cap a Heidegger en silenci, a Bares li sembla interessant el seu intent de recuperació, així com el d'altres pensadors com Pierre Hadot i Michel Foucault, de les arrels fàctiques de la comprensió en els filòsofs grecs (Aristòtil, Sòcrates o Plató).

Com esperem haver mostrat, el llibre ressenyat és capaç de portar al present i empentar a dialogar a diverses de les perspectives més fructíferes per a la teoria social contemporània. Haver assolit aquesta tasca tenint com a lligam la *història conceptual* —una perspectiva força enriquidora i amb potencialitats interdisciplinàries— i a la *crítica* —sempre necessària— és un motiu per a recomanar-ne la lectura. És cert que qui fa la ressenya troba a faltar una anàlisi que expressament problematitze la possibilitat de diàleg entre la hermenèutica (de la que prové la història conceptual) i la Teoria Crítica, dues tradicions manifestament enfrontades. Les possibilitats per a aquesta tasca estan, no obstant això, sembrades.

JOAN GALLEGO MONZÓ
Universitat de València