UNIVERSITAT DE VALÈNCIA
Facultat de Filosofia i Ciències de l'Educació
Departament de Lògica i Filosofia de la Ciencia i
Departament de Metafísica i Teoria del Coneixement

CONTRADICCIONES DE INCLUSIÓN:
LA ESTRUCTURA DE LAS PARADOJAS REFLEXIVAS Y EL PROBLEMA DE SU SOLUCIÓN

TESIS DOCTORAL

Presentada por:
Jordi Valor Abad

Dirigida por:
Jesús Alcolea Banegas y Carlos Moya Espí

Valencia, enero de 2005

UNIVERSITAT DE VALÈNCIA
FACULTAT DE FILOSOFIA I CC.EE.
Departament de Lògica i Filosofia de la Ciència
Departament de Metafísica i Teoria del Coneixement
Apt. 22.110
E-46071 VALÈNCIA - SPAIN

Jesús Alcolea Banegas y Carlos Moya Espí, directores del trabajo que presenta don **Jordi Valor Abad**, con el título ***Contradicciones de inclusión. La estructura de las paradojas reflexivas y el problema de su solución***, para la obtención del grado de doctor,

INFORMAN: Que el trabajo reúne las condiciones legales y académicas requeridas para que sea presentado y defendido ante el correspondiente tribunal.

Valencia, 2 de enero de 2005

Jesús Alcolea Banegas

Carlos Moya Espí

PRÓLOGO Y AGRADECIMIENTOS

Este trabajo se ha prolongado en el tiempo (y por ello también en la extensión que ocupan estas páginas) más de lo que habría deseado cuando lo empecé. Pese a ello, no he conseguido abordar en él muchas de las cuestiones que suscitaron desde el principio mi interés por las paradojas reflexivas, el objeto de estudio de esta investigación. Parte de las cuestiones que me preocupan (y cuyo único rastro ahora es algún comentario disperso y una nota a pie de página en el capítulo final) tendrán, pues, que ser abordadas en otro lugar. Inicialmente, los temas aquí tratados iban a ocupar las dos primeras partes de un trabajo menos extenso que tenía previsto abordar en su tercera y última parte problemas de circularidad viciosa asociados a la ejecución de ciertos proyectos semánticos y epistemológicos. Finalmente, esta última parte del proyecto original se ha abandonado por limitaciones de tiempo y espacio, sin embargo creo que el trabajo aquí desarrollado goza de suficiente unidad e independencia como para no resentirse por esta pérdida. Quisiera relatar a continuación cuáles son las principales tesis e ideas defendidas a lo largo de estas páginas, así como su contenido y estructura.

Este trabajo se divide en dos partes. En la primera (capítulos 2 y 3) se plantean dos preguntas fundamentales: ¿Es posible ofrecer una descripción estructural unificada de las paradojas? Y, si esto es posible, ¿en qué sentido puede ayudarnos dicha descripción a comprender mejor qué son y cómo se forman? La primera pregunta es respondida en el capítulo 2. Se defienden allí básicamente dos tesis:

(1) Las antinomias que Russell bautizó en 1908 con el nombre de "paradojas de autorreferencia o reflexividad" son ejemplificaciones de un mismo esquema estructural abstracto: El Esquema de Inclusión (EI). Esta estructura fue descrita en lo esencial por Russell en 1905 y ha sido defendida, bautizada y reforzada para ampliar su alcance por Priest en 1994. (El capítulo 2 contiene reducciones de varias paradojas reflexivas no tratadas por Priest y Russell a dicho esquema: la paradoja tarskiana de la cita, la del hiper-juego y la paradoja denotativa de Barwise-Moss.)

(2) Aunque el Esquema de Inclusión ofrece rasgos estructurales esenciales, i.e., condiciones *necesarias*, para describir una paradoja reflexiva, no ofrece condiciones *suficientes* y no basta, por tanto, para determinar *qué es una paradoja reflexiva*. Algo puede satisfacer EI y *no ser* una paradoja. A tal efecto, ofrecemos una reducción a dicho esquema de tres *teoremas*: el teorema de Cantor, el teorema que establece que las funciones monádicas con un solo argumento no son enumerables y el teorema que demuestra la existencia de funciones no computables por máquinas de Turing. Por ese motivo, este trabajo defiende que EI describe la forma de un determinado *tipo de razonamiento* cuya conclusión es siempre una contradicción. Que el razonamiento sea o no percibido como paradójico depende de la naturalidad con que pueda convertirse en una *reductio ad absurdum* de alguna de sus premisas. Los razonamientos que encajan en EI (sean o no paradójicos) reciben aquí el nombre de *contradicciones de inclusión*.

El capítulo 3 intenta interpretar el papel desempeñado por cada uno de los elementos de EI en la demostración de una contradicción. Se trata de ilustrar cómo los distintos componentes de EI encajan entre sí en casos típicos para alcanzar la inferencia de una contradicción. Este capítulo hace hincapié en cuatro ideas:

(1) Favorecer una determinada interpretación de los constituyentes formales de EI.

(2) Explotar las virtudes taxonómicas que ofrece EI a la hora de distinguir subfamilias dentro de las contradicciones de inclusión atendiendo sólo a criterios estructurales.

(3) Señalar la estrecha relación que guardan los razonamientos que encajan en EI con problemas filosóficos de largo recorrido histórico

(4) Señalar la consonancia existente entre EI y la mayoría de intuiciones sobre rasgos estructurales propios de las paradojas reflexivas que encontramos en la bibliografía técnica. Sobre propuestas alternativas, EI tiene la ventaja de ofrecer un análisis más general y sencillo.

La segunda parte de este trabajo parte de las respuestas que hemos dado a las anteriores preguntas y aborda, a su vez, otras dos cuestiones: ¿Qué relevancia tiene el hecho de que varias paradojas compartan una misma estructura desde el punto de vista de su solución? ¿Son satisfactorias las soluciones existentes a las paradojas de inclusión (i.e., a las paradojas que encajan en EI)? Los capítulos 4 y 5 abordan fundamentalmente la segunda pregunta y ofrecen invariablemente una respuesta negativa.

El capítulo 4 se ocupa de las soluciones a las paradojas de inclusión que buscan preservar la consistencia. Apoyándonos en el Esquema de Inclusión caracterizamos una estrategia general (a la que denominamos "parametrización") para solucionar estas paradojas de forma consistente. Así mismo, EI nos permite ofrecer una caractrerización general del dilema que acosa a todas estas soluciones: Debemos elegir entre sacrificar la consistencia de nuestras teorías o su poder expresivo y explicativo. El capítulo 4 ilustra este dilema mediante un sucinto examen crítico de las teorías consistentes más conocidas sobre la paradoja de Russell y la del mentiroso. Ninguna de ellas es plenamente satisfactoria por no poder eludir el precio que, en uno u otro sentido, nos impone el dilema descrito.

Finalmente, el capítulo 5 examina una de las aproximaciones paraconsistentes a las paradojas más serias que podemos encontrar hoy en día: el dialetheismo de Priest. Este capítulo ocupa un espacio muy superior al de cualquier otro en este trabajo por dos motivos: En primer lugar, hemos juzgado conveniente llevar a cabo un estudio bastante completo de la obra de Priest porque, pese a su seriedad y méritos innegables, su propuesta no es quizá tan conocida como debiera (i.e., goza del tipo de "fama" o "popularidad" académica que no se traduce en una aproximación real a sus obras). Muchas de las críticas recibidas por Priest incurren en peticiones de principio (al presuponer el principio de no contradicción sin ofrecer argumentos) o ignoran las respuestas y razones ofrecidas por el propio Priest a objeciones similares ofrecidas por otros o previstas por él mismo. Pese a intentar ofrecer una exposición justa de la teoría de Priest, este trabajo no comparte sus tesis teóricas. No creemos que el dialetheismo ofrezca una solución satisfactoria a las paradojas (ni siquiera una solución más satisfactoria que otras teorías rivales). Argumentamos que, al igual que sus oponentes, el dialetheista se ve forzado a introducir serias restricciones expresivas y explicativas en su entramado teórico. Si renuncia a estas restricciones, su postura desemboca fácilmente en el trivialismo (i.e., la tesis absurda de que *todo es verdadero*) y se convierte así en una propuesta inaceptable y, en cierto sentido, ininteligible.

Existe una segunda razón para prestar especial atención a las tesis dialetheistas. Priest ha afirmado que su teoría es la única que puede ofrecer una respuesta satisfactoria a la pregunta sobre la relevancia que tiene el Esquema de Inclusión para la solución de las paradojas (la pregunta que engarza ambas partes de este trabajo). Según Priest, *si todas las paradojas tienen la misma estructura, deben ser resueltas del mismo modo*. Nuestra tesis (especialmente en los capítulos 5 y 6) es que: o este principio expone una verdad general y trivial que puede ser respetada por todas las teorías (consistentes o no); o expone una falsedad, a saber, que todas las paradojas deben solucionarse *exactamente* del mismo modo. No compartimos, pues, la idea de que aceptar la contradicción que nos impone el razonamiento paradójico sea la única (o la mejor) manera de satisfacer el Principio de Solución Uniforme de las paradojas de inclusión.

A lo largo de este trabajo intentaremos afianzar una idea básica: No existe (al menos hasta el momento) una descripción estructural en términos de condiciones necesarias y *suficientes* de las paradojas (de inclusión o de otro tipo) y es probable que no exista nunca. Nuestro juicio con respecto a *qué es una paradoja* está sujeto a factores de carácter histórico que no podemos determinar algorítmicamente a partir de la presencia de ciertos rasgos sintácticos o estructurales recurrentes. Una paradoja es, de algún modo, algo que choca con "nuestra visión del mundo" y no podemos determinar *a priori* o de forma irrevisable, no histórica, cuál es "nuestra visión del mundo".

La mayoría de estudios relativos a las paradojas tienen un carácter marcadamente técnico, centrándose habitualmente en una paradoja concreta o en varias afines que tienen su origen en un mismo campo teórico. Este trabajo ha tenido que plegarse en buena medida (sobre todo en el capítulo 4) a esta circunstancia. Sin embargo, nuestro interés se centra en determinado *tipo* de paradojas, aquellas que encajan en EI, sin importarnos el campo teórico al que pertenezcan. El capítulo 1 intenta explicar (quizá no de un modo tan satisfactorio como cabría desear) por qué creemos que es relevante para la filosofía (y para la ciencia en general) el estudio de las paradojas y, en particular, el de las paradojas de inclusión. Por fortuna esta tarea ha sido recientemente llevada a cabo con gran solvencia por G. Priest en un libro continuamente presente en este trabajo: *Beyond the Limits of Thought.* Allá encontrará el lector un estudio histórico riguroso de las contradicciones de inclusión y de su relevancia en la historia de la filosofía (además de la exposición original de EI).

El capítulo 1 no forma parte relevante de esta tesis y puede ser por tanto ignorado por el lector. Salvo el capítulo 3 (que depende del 2) y el 6 (conclusiones), todos los demás gozan de cierta independencia (sólo presuponen conocer el Esquema de Inclusión) y pueden ser leídos por separado. Todos ellos (salvo el 1 y el 6) contienen un breve resumen de los capítulos anteriores en el apartado "x.1 Introducción" (el capítulo 5 está resumido en sus conclusiones, 5.4). Muchas de las notas a pie de página (sobre todo en los capítulos 4 y 5) contienen citas cuya única función es corroborar lo que se expone en el texto principal. Pueden ser igualmente ignoradas. He mantenido las citas (en vez de limitarme a hacer referencia a las obras correspondientes) por si son de utilidad para el lector. Las notas que no contienen citas o referencias bibliográficas suelen ser relevantes para la discusión de algún punto del texto principal.

Para facilitar la tarea del lector, quisiera mencionar por último algunas convenciones (referentes a la notación) recurrentes en este trabajo, aunque se harán explícitas varias veces a lo largo del mismo. En las descripciones detalladas de razonamientos formales, los pasos de inferencia aparecerán numerados formando una columna. A la derecha de cada línea de la demostración (y entre paréntesis) aparecerá su descripción o justificación (i.e., premisa, suposición, regla(s) de inferencia, etc.). La apertura y cancelación de suposiciones se representará mediante el *tabulador*. Otra convención a la que recurro a menudo (aunque no siempre) es la de utilizar el subrayado. Uso el subrayado, además, en dos sentidos diferentes (aunque no mezclo ambos en ningún ejemplo y el contexto deja claro –espero– cuál es el relevante):

A veces utilizo el subrayado como un medio para construir "contrapartidas lingüísticas" de entidades no necesariamente lingüísticas. Por ejemplo, $\underline{\mathrm{P}(a)}$ sería la "contrapartida" lingüística de P(*a*), donde el término $\underline{\mathrm{P}}$ se correspondería con la propiedad P y el término $\underline{a}$ con el objeto *a* (bajo esta perspectiva, $\underline{1}$ sería por ejemplo un *numeral* y 1 un *número*). En otras ocasiones (las más numerosas) utilizo el subrayado del mismo modo que el entrecomillado, a saber, como una forma sencilla de construir nombres de expresiones lingüísticas.

Antes de concluir quisiera señalar que este trabajo ni ofrece ni pretende ofrecer una propuesta de "solución" a las paradojas reflexivas. Todo lo que pretende es mostrar que estas paradojas guardan entre sí algo más que "cierto aire de familia" y que es bueno poner de relieve este hecho al compararlas si queremos comprender mejor cómo se producen, cuáles son sus ingredientes principales y cómo se articulan para dar lugar a una contradicción. Ahora bien, el problema que presentan las paradojas *no es formal* (aunque resulte útil apoyarnos en aspectos formales a la hora de buscar soluciones). Por sí misma, ninguna descripción estructural de las paradojas nos revela *por qué* parecen plausibles las conclusiones que nos imponen y, por tanto, ninguna descripción estructural determina cómo solucionar una paradoja. Para solucionar una paradoja debemos tomar decisiones con respecto a creencias *concretas* sobre la verdad o la falsedad de enunciados y principios que dotan de contenido a los elementos formales de nuestro análisis estructural.

Las paradojas recalcitrantes ponen de manifiesto que nuestras decisiones para atajar los problemas que plantean podrían ser sistemáticamente insatisfactorias por uno u otro motivo. Hoy por hoy, no existe una solución satisfactoria a las paradojas reflexivas que más nos preocupan, pero (pese a ser poco probable que demos con una) parece imposible establecer que no existe tal solución. Este trabajo se limita a constatar desde distintos ángulos y puntos de vista la seriedad del problema y la perplejidad que nos causa, pero no encuentra razones concluyentes o satisfactorias para decantarse por una u otra teoría. Entendemos que, en muchos contextos prácticos, debemos tomar una decisión con respecto a las paradojas, por insatisfactoria que sea (eso también forma parte del problema –pensemos en la paradoja de Russell y la teoría de conjuntos o en la del mentiroso y las semánticas formales–), pero este trabajo no constituye, afortunadamente, uno de esos contextos.

Dicen que Protágoras prometía a sus alumnos devolverles el dinero que había recibido por sus enseñanzas si perdían su primer juicio público en Atenas. Uno de ellos, Euathlo, llevó a Protágoras a juicio reclamándole el importe de su formación como abogado. Euathlo razonó así ante los jueces: "Protágoras prometió devolverme el dinero que le pagué si perdía mi primer juicio público. Ahora bien, *éste* es mi primer juicio. Si lo pierdo, Protágoras deberá devolverme mi dinero como prometió. Pero si lo gano, deberá devolvérmelo también, puesto que *ésa* es mi demanda ante ustedes." Los jueces preguntaron a Protágoras que alegaba en su defensa, Protágoras se dirigió hacia ellos y les dijo: "Euathlo, mi alumno, reclama en su primer juicio público el dinero que me pagó. Sin embargo, no veo razón alguna por la que deban acceder a su demanda, ya que, si Euathlo ganase este juicio, mi promesa no me obligaría a devolverle su dinero, puesto que *sólo* prometí devolverlo si perdía. Ahora bien, si Euathlo pierde, tampoco deberé devolverle su dinero, puesto que *ésa* es su demanda ante ustedes."

Este trabajo no envidia en absoluto la tarea encomendada al tribunal de Atenas. Por ello, no tomará partido ni por Euathlo ni por Protágoras (no sabría por quién hacerlo) y, en la medida de lo posible, evitará ocupar la incómoda posición de los jueces que, en este contexto jurídico, debieron seguramente dictar sentencia a favor de uno u otro. Desde nuestra cómoda posición de espectadores intentaremos simplemente reflejar la magnitud del problema y, eso sí, justificar las intuiciones que tenemos a favor de que muchas de estas paradojas comparten algo más que un nombre genérico y que ese hecho es relevante (aunque no determinante) a la hora de comprenderlas mejor y de abordar una estrategia de solución.

AGRADECIMIENTOS:

Quiero agradecer a la Generalitat Valenciana y a la Fundación Caja Madrid el soporte económico prestado durante los últimos seis años. Entre 1999 y 2002 fui beneficiario de una beca de formación de personal investigador concedida por la (entonces) Conselleria de Cultura, Educació i Ciència de la Generalitat Valenciana. En el año 2003, la Fundación Caja Madrid me concedió una beca predoctoral para completar mi proyecto de tesis doctoral, permitiéndome atentamente prorrogar un año más la entrega de este trabajo. Sin la ayuda financiera de ambas entidades este trabajo no se habría realizado.

Con respecto a su contenido, son muchos los reconocimientos que quisiera hacer explícitos (sobre todo teniendo en cuenta cómo se ha dilatado en el tiempo la redacción de este trabajo). Quiero en primer lugar agradecer la inestimable ayuda (y también la paciencia) de los directores de esta investigación, Jesús Alcolea Banegas y Carlos Moya Espí. Sus comentarios, anotaciones y correcciones en cada uno de los seis capítulos que contiene este estudio me han ayudado a mejorar sensiblemente su contenido y me han forzado a menudo a buscar maneras más claras y rigurosas de expresar mis ideas (aunque tal vez no siempre lo haya conseguido). Tanto como sus comentarios, quiero agradecerles sinceramente su disponibilidad a la hora de solucionar todo tipo de trámites burocráticos y su apoyo en solicitudes de becas a lo largo de estos años.

No quisiera dejar de señalar a otras personas cuyos comentarios me han sido de gran ayuda: J. Corbí y V. Sanfélix leyeron una versión preliminar del capítulo 1 y me ayudaron a discutir las ideas allá expuestas, simplificar su contenido y eliminar algunos errores. Diferentes versiones del capítulo 2 (y secciones del 3) fueron presentadas en dos congresos: "Symposium on Alfred Tarski and the Foundations of Logic and Semantics" (Instituto de Filosofia da Linguage, Universidade Nova de Lisboa. Lisboa octubre de 2003) y el "VIè Taller d'Investigació en Filosofia" (Universitat Rovira i Virgili. Tarragona, enero de 2004). Quiero agradecer en general a los asistentes de ambos congresos sus comentarios, sugerencias y rectificaciones (en particular a P. Cobreros, J. Corcoran, J. Díez, M. Gómez, D. López de Sa, J. Macià, J. M. Sagüillo, y R. Santos). El contenido y las ideas del capítulo 4 es el que más ocasiones he tenido de discutir y contrastar. Muy especialmente durante una breve estancia en la Universidad de Sheffield en la que tuve la ocasión de trabajar sobre teorías de la verdad en un grupo de lectura muy animado y en diversos seminarios en el Departamento de Metafísica de Valencia. Quiero destacar en este apartado a: J. Martínez y R. Keefe. También a C. Abell, M. Bares, O. Benito, M. Campdelacreu, T. Grimaltos, C. Hookway, D. Kuenzle, P. Martínez, M. Moreno, J. Saul, A. Sherrat, V. Vidal e I. White. Con respecto al capítulo 5, quiero agradecer encarecidamente las valiosas correcciones y comentarios de J. P. Úbeda, su lectura de este capítulo fue de gran utilidad para mí.

Junto a los directores de este trabajo, José Martínez Fernández es sin duda la persona con quien más he discutido y pensado sobre los temas que aquí trato. Quiero destacar la inestimable ayuda que me ofreció al principio de mi investigación (cuando más necesaria era) a la hora de seleccionar bibliografía y "digerir" mejor algunas de las ideas y "oscuros" tecnicismos formales que acompañan frecuentemente las discusiones sobre paradojas reflexivas en las revistas especializadas.

Quiero agradecer así mismo el apoyo recibido en general por los miembros de los Departamentos de Lógica y Filosofía de la Ciencia y Metafísica y Teoría del Conocimiento, y, en particular, a los miembros del PAS: A. Castellano, A. Garriga, A. Gómez, A. López y E. Valero (por haber evitado en más de una ocasión que me perdiese en uno de esos turbios laberintos burocráticos que me "aterran" mucho más que la peor de las paradojas).

Por último, aunque ciertamente no en último lugar, quiero agradecer el apoyo de todos mis amigos (entre los que cuento a muchos de los ya mencionados y, especialmente, a O. Soler y M. Oliver). Pero, sobre todo, quiero agradecer el apoyo de mi familia, mis padres, Maria Lluïsa y Jordi, y mi hermano, Jero, que son junto a mí quienes más alteradas han visto sus vidas por culpa de los años que me ha "robado" (a veces no muy gustosamente) este trabajo. Les agradezco a todos su cariño, su apoyo incondicional y su bendita paciencia conmigo y con esta "cosa" que ya por fin he terminado (espero).

ÍNDICE

CONTRADICCIONES DE INCLUSIÓN:

La estructura de las paradojas reflexivas y el problema de su solución

1. La relevancia filosófica de las paradojas..1

1.1 Paradojas: Algunas cuestiones preliminares..1

1.2 El problema..7

1.3. Dos imágenes de las paradojas (I): paradojas y progreso cognitivo...................25

1.4. Dos imágenes de las paradojas (II): paradojas y límites cognitivos....................35

I EL ESQUEMA DE INCLUSIÓN Y LAS PARADOJAS REFLEXIVAS..........51

2 La estructura de las paradojas reflexivas: el Esquema de Inclusión................53

2.1 Introducción..53

2.2 La clasificación de las paradojas "reflexivas": Russell contra Ramsey................55

2.3 ¿Cuál es el problema?...61

2.4 El esquema de Russell..69

2.5 El Esquema de Inclusión [*Inclosure Scheme*]..73

2.6 La "paradoja" de la cita de Tarski..79

2.7 Razonamientos que satisfacen el Esquema de Inclusión: el teorema de Cantor 87

2.8 Críticas al Esquema de Inclusión..92

2.8.1 Las paradojas de Yablo y Curry no encajan en EI...92

2.8.2 ¿Define EI las paradojas en términos de condiciones necesarias y suficientes?.96

2.8.3 Algunos elementos de EI parecen redundantes en algunas paradojas.............98

2.8.4 La clasificación de Ramsey subsiste...100

2.8.5 Con algún artificio, toda contradicción encaja en EI.....................................101

2.9 Conclusiones..102

2.10 Contradicciones de inclusión. (Apéndice capítulo 2)..102

2.10.1 La paradoja de la denotación de Barwise y Moss..103

2.10.2 La paradoja de Yablo..105

2.10.3 La paradoja del cocodrilo..110

2.10.4 La paradoja del hiper-juego..113

2.10.5 Funciones no enumerables y funciones no (Turing-)computables..115

2.10.6 La paradoja de Curry..119

3 Cómo interpretar EI: diagonalizadores y totalidades absolutas..122

3.1 Introducción..122

3.2 ¿Qué significado tiene el Esquema de Inclusión? Los límites del pensamiento.....124

3.2.1 Generadores iterativos, ω-secuencias y secuencias transfinitas..124

3.2.2 Límites iterativos y totalidades absolutas..127

3.2.3 Iteración y orden..130

3.2.4 La serie de los ordinales, On..133

3.2.5 El Esquema de Inclusión, cómo interpretar δ: diagonalizadores..140

3.2.6 El Esquema de Inclusión, cómo interpretar Ω y ψ: el alcance de δ..143

3.2.7 Contradicciones en los límites de la iteración: Burali-Forti..148

3.2.8 ¿Contradicciones en los límites de la iteración?: Russell..153

3.2.9 Contradicciones en las que ψ no es una propiedad universal..157

3.2.10 Contradicciones en las que ψ sólo es satisfecha por Ω..163

3.2.11 Otros criterios estructurales para clasificar a las contradicciones de inclusión........165

3.2.12 ¿Cuál es el problema de fondo en las contradicciones de inclusión?..170

3.2.13 Un problema filosófico..172

3.3 Algunas intuiciones sobre la estructura de las paradojas reflexivas y EI...........180

3.3.1 Autorreferencia, reflexividad, circularidad y EI..180

3.3.2 Negación y diagonalización..183

II SOLUCIONES A LAS PARADOJAS DE INCLUSIÓN............................189

4 El Principio de Solución Única, enfoques consistentes y "parametrización"....191

4.1 Introducción...191

4.2 Condiciones para una solución satisfactoria de las paradojas reflexivas: el PSU.. 193

4.3 Diagonalización y estrategias de solución "paramétricas" (ejemplos)...............200

4.3.1 Estrategias "paramétricas": Restricciones de Ω y δ.................................203

4.3.2 El problema de la diagonalización: Paradojas reforzadas.........................215

4.4 Soluciones a las paradojas conjuntistas..220

4.4.1 ZF: Zermelo y la Jerarquía Acumulativa de Conjuntos..............................221

4.4.2 Von Neumann y las clases propias..227

4.4.3 Conclusión...230

4.5 Soluciones a las paradojas "semánticas" (la paradoja del mentiroso)................231

4.5.1 Russell y Tarski: Las jerarquías de tipos y lenguajes..............................231

4.5.2 Lógicas no clásicas, fallos semánticos y puntos fijos..............................237

4.5.3 Teorías contextualistas: verdadero en un context..................................246

4.5.4 Teorías revisionistas con bucles cíclicos (Gupta y Herzerberger)..................265

4.6 Conclusiones generales..272

5 El Dialetheismo de Priest...277

5.1 Introducción...277

5.2 El Dialetheismo de Priest frente a otras teorías: Coherencia, generalidad y PSU...279

5.2.1 Indicios o pruebas a favor del Dialetheismo de Priest..............................283

5.2.1.1 El teorema de Gödel: Teorías inconsistentes de la demostración...286

5.2.2 ¿Es coherente el dialetheismo y racional aceptar una contradicción?.............292

5.2.2.1 Principios de No Contradicción y Ex Contradictione Quodlibet....296

5.2.2.2 Verdadero, no verdadero, falso...305

5.2.2.3 La negación...310

5.2.3 Lógica paraconsistente (LP) y modelos inconsistentes.............................315

5.2.3.1 La semántica de LP...315

*5.2.3.2 LP es mínimamente inconsistente: razonamientos "cuasi-válidos"*323

5.2.3.3 Modelos inconsistentes......................................326

5.3 Críticas al dialetheismo......................................338

5.3.1 Ausencia de criterios sintácticos para detectar dialetheias......................342

5.3.2 Restricciones inferenciales......................................344

5.3.3 La idea de absurdo y el significado de negar: Negación y rechazo (I)..........346

5.3.4 Falso y no verdadero: Negación y rechazo (II)......................353

5.3.5 Incoherencias y paradojas reforzadas propias del dialetheismo..............372

5.3.6 Críticas al Principio de Solución Uniforme (PSU)......................392

5.3.7 El Teorema de Gödel no muestra la inconsistencia de nuestro razonamiento..401

5.4 Conclusión: ¿Es el dialetheismo una solución satisfactoria a las paradojas?..........406

6 Conclusiones......................................413

6.1 Recapitulación: El valor real del Esquema de Inclusión......................413

6.2 De nuevo el viejo problema......................................428

Apéndice (Índice de paradojas de inclusión)......................................435

Bibliografía......................................445

1. La relevancia filosófica de las paradojas

1.1 Paradojas: Algunas cuestiones preliminares

¿Qué entendemos habitualmente por "paradoja"? Quizá sea ésta la pregunta que hayamos de formularnos antes de examinar el objeto de este capítulo inicial: qué relación guardan las paradojas con la filosofía. Ciertamente éste sería el camino a seguir para filósofos que, como Wittgenstein o J. L. Austin, pensaron que los problemas de la filosofía a la hora de comprender el significado de ciertos términos lingüísticos provienen en su mayoría de una negligencia. A saber, la de ignorar de forma sistemática –ya sea por descuido o deliberadamente– el uso habitual que dichos términos tienen en las prácticas a las que pertenecen originalmente, aquellas que determinan su significado genuino. Seguramente ni Wittgenstein ni Austin pretendían exhortarnos a resolver problemas filosóficos "acudiendo al diccionario" (aunque tal vez Austin habría agradecido que lo consultásemos más a menudo). Los problemas filosóficos no se resuelven así, pero es bueno en todo caso que nuestras reflexiones comiencen al menos en un lugar en el que no se dé por sentado que el sentido de ciertos términos está determinado por el uso que de ellos hacemos en la práctica de la filosofía académica.

Desde esta perspectiva, un diccionario puede ser un punto de partida tan válido como cualquier otro para abordar la anterior pregunta (aunque no deje de ser eso, un punto de partida y, desde luego, no el único). Después de todo, el trabajo que allí desarrolla un lingüista es sobre todo empírico y busca precisamente captar el uso que una comunidad de hablantes hace de los términos de un lenguaje –aunque toda inclusión o exclusión tenga también intenciones normativas–. Si acudimos a un diccionario o a una enciclopedia,[1] descubriremos que existe un amplio consenso a la hora de primar como acepción fundamental del término 'paradoja' una descripción muy ceñida a su origen etimológico griego ('παραδοξος'). Según esta acepción, toda paradoja es "un enunciado o una idea contraria a la opinión común de las personas, algo que nos parece absurdo, increíble o inaceptable" (DRAE, DIEC, DUO). En un sentido más técnico o específico, las paradojas se asocian al campo de la lógica y las matemáticas y se describen como "razonamientos aparentemente válidos que, partiendo de premisas verdaderas, alcanzan conclusiones falsas" (DIEC, DUO).[2] En estos casos es frecuente (aunque no necesario) que la conclusión tenga la forma de una contradicción, 'A y no A', y que el razonamiento ofrezca un argumento a favor de 'A' y otro a favor de 'no A'. Pese a ser tildada de "técnica" o "específica", esta acepción –habitual en los diccionarios, pero curiosamente ausente en DRAE– añade un elemento que está presente en todas las paradojas y en nuestra concepción ordinaria de las mismas. Una

---

[1] En el apéndice final (que contiene un índice de paradojas) incluimos las entradas que para el término 'paradoja' aportan diccionarios de la Real Academia de la Lengua Española (DRAE), del Institut d'Estudis Catalans (DIEC) y de la Universidad de Oxford (DUO).

[2] Existe una cierta ambigüedad a la hora de determinar si la paradoja es el *razonamiento* aparentemente válido que conduce a una conclusión falsa, el enunciado falso de la conclusión (en tanto que conclusión de un razonamiento paradójico) o cualquier enunciado (aparentemente verdadero) que desencadene un razonamiento válido con

paradoja es algo que *tiene su origen en un razonamiento.* Sin razonamiento no hay paradoja. Una mera falsedad como "África es un país", una contradicción como "tengo sueño y no tengo sueño" o un absurdo como "2 = 1" *no son* paradojas (pese a contradecir claramente nuestra opinión), las paradojas surgen cuando somos capaces de *justificar mediante un razonamiento* la verdad, la validez o el sentido de enunciados o ideas aparentemente falsas, contradictorias o absurdas, esto es, cuando tenemos una justificación racional para aceptar lo que parece inaceptable. Enunciados como "no existe el cambio" o "el conocimiento es imposible" no son paradójicos en sí mismos, pese a chocar plenamente con nuestras convicciones al respecto. Sin embargo, si construyésemos un argumento convincente a favor de estas conclusiones, un argumento que apelase a principios válidos y partiese de premisas plausibles (como hicieron en un caso Zenón y en el otro Sexto Empírico, por ejemplo), entonces tendríamos un problema. En la medida en que el razonamiento nos pareciera válido y sus premisas verdaderas, *tendríamos razones* para aceptar cosas que juzgamos *falsas* o *absurdas.*

Aunque no es necesario que una paradoja se presente como un razonamiento contradictorio, cualquier argumento que dé origen a una paradoja podría adoptar esta forma. La razón es sencilla: *Todo* razonamiento aceptable que parta de premisas verdaderas ha de establecer enunciados *verdaderos.* Así pues, la existencia de un razonamiento *válido,* R, con premisas *verdaderas* y conclusión C

---

conclusión falsa. Lo más justo quizá sería identificar la paradoja con el razonamiento, en cualquier caso ésa será la opción a seguir en este trabajo.

constituye en principio una prueba o un indicio a favor de *la verdad* de C, no de su falsedad. Decir que C *es falso* es sugerir que tenemos *motivos* o *razones* para justificar la falsedad de C *con independencia de* R. Ahora bien, esos motivos o razones se pueden articular en forma de un razonamiento R' (no necesariamente complejo) a favor de la verdad de la *negación* de C. De ese modo dispondríamos de un razonamiento, R, a favor de "C" y de otro, R', a favor de "no C". Muchas veces se opta por describir la paradoja como el resultado de unir ambos argumentos y sus conclusiones contradictorias.

Otro rasgo común a las paradojas es que la conclusión que nos imponen provoca en nosotros un rechazo inmediato, espontáneo. Para enfatizar esta idea muchos diccionarios afirman que las paradojas van contra el "*sentir* común de las personas" (DIEC, DRAE). La conclusión de un razonamiento paradójico nos parece *obviamente falsa*, lo no tan obvio, lo sorprendente, es el hallazgo de un razonamiento válido con premisas verdaderas que justifique algo tan claramente opuesto a nuestras creencias. Este rasgo de las paradojas también influye en nuestra decisión de presentarlas como argumentos que desembocan en una contradicción o en una mera falsedad. Si la falsedad es muy evidente (por atentar contra la información que nos aportan los sentidos –como ocurre con las paradojas de Zenón–, o contra leyes lógicas básicas, etc.), entonces la paradoja no precisa formularse como la justificación de una contradicción. Si la falsedad es más sutil, entonces la paradoja suele expresarse buscando una conclusión contradictoria ya que toda contradicción es percibida siempre de forma inmediata como algo rechazable y absurdo.

Resulta obvio, a tenor de lo dicho hasta ahora, que enfrentarse a una paradoja es enfrentarse a un serio problema. Un razonamiento válido con premisas verdaderas *no puede* tener una conclusión falsa. Por este motivo la mayoría de textos centrados en las paradojas –incluidas las definiciones de un diccionario– nos previenen contra la inmediatez de nuestros juicios iniciales sobre ellas y nos advierten que *las apariencias engañan*. Si no existen razonamientos válidos con premisas verdaderas y conclusión falsa, entonces las paradojas son el producto de un error de apreciación. Tal vez alguna de las premisas no sea verdadera, en contra de lo que pensábamos; tal vez la conclusión no sea falsa después de todo; o quizá las leyes de inferencia aplicadas en nuestro razonamiento no sean válidas (o su aplicación pertinente), como creíamos. Algo tiene que fallar en nuestro juicio sobre la verdad de las premisas, la falsedad de la conclusión o la validez del razonamiento. La mayoría de diccionarios recoge esta idea en alguna de sus acepciones y hablan así de razonamientos "aparentemente" válidos, de conclusiones "aparentemente" absurdas o falsas, o de presupuestos dudosos. Frente a una paradoja se tiene siempre la esperanza de reestablecer el orden roto revisando alguna de nuestras impresiones sobre el razonamiento que la produjo.

Tal vez esta esperanza de que toda paradoja sea en última instancia un reto intelectual, un acertijo o un laberinto "con salida", sea la responsable de que el término 'paradoja' se asocie también (como refleja cualquier diccionario) a ciertos recursos retóricos o literarios consistentes en presentar ideas *verdaderas* y perfectamente *significativas* bajo *la apariencia* de enunciados sin sentido o contradictorios.[3]

Pese a todo, también en los diccionarios encontramos a veces el temor o incluso la sospecha de que algunas paradojas no sean resolubles de modo satisfactorio o, al menos, de un modo que elimine la perplejidad que nos causan (de los citados, DIEC es quizá el que deja traslucir mejor esta posibilidad). Se podría decir que resolver una paradoja es en cierto modo *disolverla*, *destruir la ilusión* que nos ha llevado a creer que el razonamiento paradójico es válido, o sus premisas verdaderas, o su conclusión falsa. Una vez la ilusión es expuesta como tal y reconocida por todos, la ilusión se desvanece, como ocurre cuando descubrimos que el mago no ha hecho desaparecer la moneda, sino que la oculta entre sus dedos. En ese momento la paradoja deja de ser percibida como tal (i.e., como algo "extraño o ajeno a la opinión y el sentir común de las personas") y pasa a ser simplemente un argumento erróneo o mal construido. Contemplar la posibilidad de que algunas paradojas no sean "ilusiones", es contemplar la posibilidad de que existan paradojas "genuinas" –por llamarlas de algún modo– que nunca podremos domesticar o disolver.

[3] Los famosos versos de San Juan de la Cruz en las *Coplas del alma que pena por ver a Dios* son un ejemplo clásico: "Vivo sin vivir en mí// y de tal manera espero,// que muero porque no muero.// En mí yo no vivo ya,// y sin Dios vivir no puedo;// pues sin él y sin

1.2 El problema

Está claro a partir de las observaciones que acabamos de hacer cuál es el reto o problema fundamental que presenta cualquier paradoja. Una paradoja nos enfrenta a la posibilidad de aceptar razonamientos que partiendo de premisas verdaderas alcanzan conclusiones falsas, contradictorias o absurdas. Esta posibilidad choca de frente con nuestra idea de lo que supone razonar correctamente, así pues, debemos reaccionar frente a ella de algún modo. Para apreciar la seriedad del problema será conveniente que nos detengamos un momento en sus consecuencias.

El objetivo esencial de razonar es descubrir qué información nueva podemos obtener a partir de lo que ya sabemos. Cuando razonamos a partir de ciertos enunciados (premisas) que nos parecen verdaderos, nuestra meta es encontrar enunciados cuya verdad pueda justificarse a partir de la verdad de los enunciados de partida. Un razonamiento concluye cuando conseguimos justificar correctamente (aplicando reglas de inferencia válidas) la verdad de los nuevos enunciados a partir de la verdad de las premisas. Si los razonamientos válidos no preservasen la verdad de las premisas en la conclusión, razonar no resultaría una actividad cognitiva *fiable* y perdería por tanto su función esencial, proporcionar conocimiento. Es cierto que algunas formas de razonamiento, como por ejemplo la inducción empírica, son útiles pese a ser *falibles*, esto es así porque habitualmente nos ayudan a establecer

---

mí quedo,// este vivir, ¿qué será?// Mil muertes se me hará,// pues mi misma vida espero// muriendo, porque no muero."

enunciados *verdaderos*. Pero es importante señalar que *no es lícito* aceptar un razonamiento inductivo sabiendo que establece una conclusión *falsa* y la razón de ello es, precisamente, que el objetivo de todo argumento aceptable es establecer verdades. Como las paradojas incumplen esta máxima, deben ser rechazadas. Con todo, estas observaciones no permiten vislumbrar aún la gravedad real del problema.

Las lógicas consistentes (con mucha diferencia, las dominantes) aceptan el principio de no contradicción y, junto a éste, aceptan también el de "*ex contradictione quodlibet*" según el cual de una contradicción se sigue *cualquier cosa*. Hemos dicho que toda paradoja se puede expresar mediante un razonamiento cuya conclusión es una contradicción. Si aceptásemos la validez de un razonamiento paradójico, podríamos establecer una contradicción y a partir de ella *cualquier cosa*, lo que implicaría en la práctica la inutilidad del razonamiento lógico. Si pudiésemos demostrar *cualquier cosa* razonando, si *todo* fuese demostrable, la actividad de razonar se convertiría en algo irrelevante y trivial ya que ningún razonamiento ayudaría a separar aquellos enunciados cuyo contenido deseamos aceptar de aquellos que deseamos rechazar. Si un razonamiento fuese además algo que justifica la verdad de un enunciado, como hemos señalado, entonces *todo* enunciado resultaría verdadero y perdería así sentido cualquier práctica cognitiva destinada a distinguir la verdad de la falsedad, lo que convertiría a la filosofía, la semántica, la lógica y, en general, a cualquier ciencia formal o empírica en actividades inútiles.

Estas consecuencias son lo suficientemente dramáticas como para que muchos autores se tomen en serio (y no como un mero juego intelectual) el problema que cualquier paradoja supone por peculiar, rebuscada o intrascendente que parezca para nuestra vida cotidiana.[4] Este catastrófico panorama explica también que la inmensa mayoría de filósofos, matemáticos, lógicos, etc. apoyen lo que podríamos llamar una "*visión doxástica*" de las paradojas (preponderante también en los diccionarios). Según esta visión, una paradoja es "un razonamiento *aparentemente* válido que partiendo de premisas *aparentemente* verdaderas, establece una conclusión *aparentemente* falsa". Esta concepción destierra por completo la posibilidad de un razonamiento válido con premisas verdaderas y conclusión falsa. Nuestra creencia en la validez de una paradoja no puede constituir conocimiento, sólo puede obedecer a *opiniones* erróneas con respecto a alguno de los elementos involucrados en el razonamiento. Nuestra misión es descubrir dónde yace el error (premisas, conclusión o razonamiento), explicar nuestro engaño y deshacer la ilusión de la que somos víctimas.[5]

[4] Tarski expresa vívamente su sensibilidad hacia este problema: "A mi juicio, sería erróneo y peligroso, desde el punto de vista del progreso científico, despreciar la importancia de ésta [la paradoja del mentiroso] y otras antinomias, tratándolas como bromas o sofistiquerías. Es un hecho que estamos en presencia de un absurdo, que nos hemos visto obligados a afirmar una oración falsa [la oración *s* (que designa a '*s* no es verdadera')]. Si tomamos en serio nuestro trabajo, no podemos ignorar este hecho. Debemos descubrir su causa, es decir, debemos analizar las premisas sobre las que se basa la antinomia; luego debemos rechazar por lo menos una de esas premisas, y debemos investigar las consecuencias que esto tiene para el dominio íntegro de nuestra investigación." Tarski 1944, VII, p. 284.

[5] "This is what I understand by a paradox: an apparently unacceptable conclusion derived by apparently acceptable reasoning from apparently acceptable premises. Appearances have to deceive, since the acceptable cannot lead by acceptable steps to the

Frente a la visión doxástica de las paradojas encontramos también otra con menos apoyos, aunque no por ello carente de seguidores. A falta de un nombre mejor, llamaremos a esta concepción de las paradojas "*óntica*". Según esta visión del problema, las paradojas nos muestran que ciertas contradicciones son inevitables, que existen de hecho en el mundo y podemos establecer su verdad. Los defensores de esta postura son mucho menos numerosos que sus oponentes, pero entre sus filas podríamos encontrar a filósofos como Hegel y quizá, remontándonos aún más en la historia, Heráclito.[6] Quienes favorecen la visión óntica de las paradojas hacen hincapié en el hecho de que algunas paradojas nos han acompañado durante siglos sin ser resueltas de modo satisfactorio y añaden además que no existen indicios de que esta situación vaya a cambiar en el futuro, sino más bien de lo contrario.

Las razones que aportan los defensores de la visión doxástica a favor de su postura no dependen de cuestiones empíricas, son previas al

unacceptable. So, generally, we have a choice: either the conclusion is not really unacceptable, or else the starting point, or the reasoning, has some non-obvious flaw." Sainsbury 1995, p. 1. Son incontables las citas parecidas de diversos autores que podemos encontrar en la bibliografía técnica. Esta visión está implícita, sin ir más lejos, en la cita de Tarski de la nota anterior.

[6] Hoy en día, quizá sea Priest el principal lógico defensor de la visión óntica: "I claim that reality is, in a certain sense, contradictory. I do not, of course, mean that the objects that constitute reality, like chairs and stars, are contradictory. That would simply be a category mistake. What I mean is that there are certain contradictory statements (propositions, sentences –take your pick) about limits, that are true. I am enough of a realist to hold that there must be something about reality that makes them so [...] When I say that reality is contradictory, I mean that it is such as to render those contradictory statements true. If we are to think about that reality in an adequate fashion, it follows that those contradictions must be part of the content of our thought." (G. Priest 2002, p. 295). Priest matiza además que "[t]he concern with thought here is not with subjective mental states, but with objective content; and one cannot divorce this content of thought from the reality of which the thought is about." (Ibídem, p. 294)

dictamen de la experiencia: La idea de un razonamiento válido con premisas verdaderas y conclusión contradictoria (una paradoja) es *absurda* e inaceptable *a priori*, rompe las leyes del razonamiento. Por su parte, los defensores de la visión óntica recurren a argumentos *a posteriori* en los que, apelando a la experiencia, se insiste en que la aplicación de nuestros principios racionales a la solución de las paradojas ha fracasado sistemáticamente, no ha conseguido eliminar las contradicciones que éstas nos imponen. Tenemos que ver pues cómo es posible acomodarlas y darles sentido. Se establece así una dialéctica entre dos maneras antagónicas de responder a las paradojas.

Aunque las actitudes de los defensores de una y otra concepción son claramente opuestas, la línea que divide ambos bandos no es siempre nítida. Habitualmente resulta difícil mantenerse de forma coherente o satisfactoria en uno de los extremos. Ambos bandos reconocen el peso y la fuerza de los argumentos de su rival. Quienes tratan de ofrecer una solución consistente a paradojas tan persistentes como la del mentiroso (el enunciado que dice de sí mismo que es falso) han tropezado siempre con la evidencia empírica de que sus teorías aportan los elementos necesarios para formular versiones reforzadas o más sofisticadas de la paradoja que pretendían resolver. Aceptar la verdad de esas teorías y las consecuencias que de ellas se siguen no difiere mucho por tanto de aceptar contradicciones (las que implica la propia teoría). Por su parte, los defensores de la visión óntica son conscientes de que las motivaciones *a priori* de sus oponentes son poderosas y se apropian en parte de ellas: Razonar es buscar conocimiento y sólo podemos conocer *lo verdadero.* Si aceptásemos razonamientos que no preservasen la verdad de las premisas en la conclusión, razonar no nos ayudaría a conocer. Para salvar este escollo los partidarios de teorías paraconsistentes reivindican "*la verdad*" de ciertas contradicciones (aquellas que establecen las paradojas) y rechazan el principio de *ex contradictione quodlibet.* De ese modo, la verdad de las premisas se preserva siempre en la conclusión de un razonamiento válido (aunque éste sea paradójico) y se evita la inferencia de enunciados arbitrarios. Por desgracia, es un hecho empíricamente contrastable que los defensores de este punto de vista también son víctimas de paradojas reforzadas. En su caso las paradojas suelen reintroducir la posibilidad absurda de que todo sea verdadero, el trivialismo.

Así pues, junto al problema fundamental (no empírico) que suponen las paradojas –la posibilidad de un razonamiento válido que *no preserve la verdad* de las premisas en la conclusión–, tropezamos ahora con un problema empírico relacionado con nuestras respuestas al problema fundamental: algunas paradojas (ya milenarias) no han encontrado hasta la fecha una solución satisfactoria y nada indica que esta situación vaya a cambiar en un futuro razonablemente cercano (o incluso lejano). Todo intento de solución ha generado nuevas paradojas y perplejidades sin importar que el enfoque de partida fuese "doxástico" u "óntico". La constatación de este hecho justifica que nos planteemos cuál es la significación real de las paradojas y cómo afectan éstas a nuestra concepción general del mundo y del conocimiento. Muchas de las ideas que asociamos con las paradojas en este sentido obedecen a convicciones y certezas de carácter metafísico que, pese a buscar apoyos en nuestra experiencia, no podrían nunca ser contrastadas o verificadas empíricamente. En el mejor de los casos adoptan la forma de argumentos trascendentales en los que se justifica la posición adoptada como la única manera de hacer inteligible el significado que las paradojas tienen en nuestra vida.

El origen de la imagen metafísica de las paradojas podría describirse a grandes rasgos de la siguiente manera: Razonar es una actividad encaminada a obtener conocimiento y por tanto a establecer verdades. Cualquier paradoja nos enfrenta a la posibilidad de conceder validez a razonamientos inaceptables cuyo objetivo no es la verdad. Caben dos lecturas de este hecho: O la paradoja revela que nuestras convicciones y principios teóricos son erróneos y puede entonces resolverse revisando dichas convicciones y principios; o la paradoja no tiene solución y muestra por contra que el razonamiento humano no siempre es una herramienta eficaz para obtener conocimiento (ya que genera absurdos bajo ciertas circunstancias). Esta disyuntiva puede desembocar en un callejón sin salida por los siguientes motivos: (1) Solucionar una paradoja requiere siempre una investigación exhaustiva,[7] lo que significa que carecemos de justificación para asumir que una paradoja no tiene solución por el mero hecho de no haber podido solucionarla. Siempre podemos imaginar que nuestros esfuerzos en esa

[7] Tal vez la investigación no tenga un carácter empírico, tal vez apele a factores independientes de la experiencia, pero incluso un teorema debe ser demostrado tanto para saber que lo es como para ser aceptado. Así pues, dar una respuesta a esta pregunta no es nunca una tarea trivial.

dirección han sido insuficientes. (2) Por otra parte, ninguna investigación podría establecer que una paradoja carece de solución, para ello tendríamos que construir un argumento que demostrase que el razonamiento paradójico es *válido* pese a *no preservar* la verdad, pero ese argumento sería en sí mismo inaceptable, ya que atentaría contra nuestra concepción de lo que significa razonar. En otras palabras, ese argumento constituiría un absurdo, una paradoja sobre la que cabría preguntarse si tiene o no solución. (3) Esto supone que la cuestión de si una paradoja es o no "genuina" (entendiendo por tal irresoluble) podría ser *indecidible* en los casos problemáticos. Si nos hallásemos en presencia de una paradoja genuina, nunca podríamos determinar con certeza absoluta que es irresoluble, de donde inferimos que *no es posible saber* si existen paradojas genuinas. (4) Ahora bien, nuestra ignorancia *no demuestra* que no haya paradojas genuinas, lo que nos impide excluir por completo esa posibilidad en nuestras hipótesis teóricas, pese a resultarnos absurda. Esta incómoda situación da origen a dos poderosas imágenes metafísicas de la significación que tienen las paradojas en nuestra vida cognitiva. Según una imagen, las paradojas tienen un carácter "imperativo". En la medida en que plantea la posibilidad de un absurdo inaceptable, toda paradoja nos impone la obligación de encontrar una solución. Desde este punto de vista las paradojas son el motor del progreso cognitivo. Según la otra imagen la posible existencia de paradojas genuinas constituiría una manifestación de los límites del conocimiento humano y de nuestras capacidades cognitivas. Si pudiésemos identificar una paradoja genuina, dotaríamos de contenido concreto la idea de límite cognitivo, podríamos "ver" de alguna manera cuál es la raya que no podemos traspasar.

La primera imagen de las paradojas hace que éstas desempeñen un papel esencial en el desarrollo de las ciencias, son en cierto sentido el motor del conocimiento. Tal vez no hayamos resuelto "aún" todas las paradojas, pero no tenemos motivos para dejar de intentarlo y sí en cambio para perseverar. Todo intento de solución ha generado progresos notables en nuestra visión del mundo y ha perfeccionado nuestros mecanismos para representarlo. Es un hecho que la forma que tenemos de afrontar las paradojas y los conceptos que empleamos en su solución y formulación han ido evolucionando para adaptarse a nuestros fracasos y ofrecer así mejores respuestas. Muchas paradojas, por ejemplo las relativas al infinito, no se formulan hoy (Burali-Forti [7], Mirimanoff [9]) en los términos en los que se formulaban en tiempos de Aristóteles o de Zenón. Algunas antiguas paradojas ni siquiera son percibidas hoy en día como tales (pese a esconder problemas genuinos si se piensan detenidamente) debido a la extrañeza que nos produce actualmente la supuesta verdad de sus premisas o el sentido conferido a los términos que la sustentan. El rechazo que provoca toda contradicción demostrada a partir de principios aceptados estimula la vida y la evolución histórica de cualquier ciencia ya sea formal, natural o social. Dado que no tenemos experiencia de una ciencia general completamente libre de contradicciones (i.e., irrefutable e irrevisable) y tampoco podemos comprender que una teoría contradictoria sea aceptable, la contradicción se percibe como un "mandamiento" (casi moral) que nos insta a buscar el progreso incansablemente. De nuestro intento por cumplir ese mandato surge la historia de toda disciplina científica.[8]

Pero junto a este imperativo ineludible, convivimos con el temor (que aviva la segunda imagen metafísica de las paradojas) de encontrarnos en la situación de modernos Sísifos, condenados a empujar la *misma* piedra todos los días a la cima de la *misma* montaña. Si algunas paradojas no tuviesen solución, tratar de solucionarlas *nos alejaría de la verdad* en vez de acercarnos a ella. Obtener conocimiento acerca del significado que estas paradojas tienen en nuestra vida cognitiva equivaldría a *demostrar* que son irresolubles, sólo así podríamos conocer la verdad que encierran. El problema radica, como ya hemos señalado, en que, de existir, "esa verdad" es inaccesible para nosotros. Trazar los *límites absolutos* del conocimiento humano supone ubicarse en una posición que nos está vedada. Deberíamos ser capaces de identificar una línea divisoria *concreta* y "*tangible*" que separase lo que conocemos de lo que *no podemos* conocer, pero eso es imposible. Tal vez fuésemos capaces de identificar la línea que separa las cosas que conocemos de las que ignoramos (y eso en sí ya es dudoso), pero para saber que las cosas que ignoramos *no pueden ser conocidas* deberíamos ser capaces de asomarnos al otro lado de la línea y "ver que no es posible ver", ahora bien, eso es precisamente lo que carece de sentido. Si podemos romper la

[8] Son muchos los filósofos que, desde Kant y (sobre todo) Hegel, han alimentado este tipo de imágenes "vitales" haciendo que trascendieran al campo del pensamiento científico e incluso al del pensamiento político y económico. Sirva como ejemplo esta cita de Engels utilizada por Priest (1987, p. 277): "...and as soon as contradiction ceases, life, too, comes to an end and death steps in." (La referencia de Priest es: Engels, F., *Anti-Dhüring*, 1894, 3ª edición (traducción inglesa 1975, Progress Publishers, p. 140)).

línea y ver lo que hay al otro lado, esa línea no es el límite absoluto que buscamos, necesitamos trazar una nueva. Si, por contra, no podemos romper la línea, ni por tanto contemplar lo que hay al otro lado, todo lo que legítimamente estamos autorizados a inferir es que *no hemos roto la línea*, no que la línea sea *irrompible*.[9] Si no fuésemos humanos y conociésemos todo lo que conocen los humanos y más cosas, entonces tendría sentido para nosotros la empresa de delimitar los límites del conocimiento humano con respecto al nuestro (aunque no los límites de

---

[9] Una idea de este tipo aparece perfectamente prefigurada en el *Tractatus* cuando Wittgenstein habla de la posición que el sujeto metafísico ocupa en "su" mundo comparándola con la que ocupa el ojo con respecto a su campo visual (5.631-5.6331):
"5.632 El sujeto no pertenece al mundo, sino que es un límite del mundo.
5.633 ¿Dónde descubrir en el mundo un sujeto metafísico?
Dices que ocurre aquí enteramente como con el ojo y el campo visual. Pero el ojo *no* lo ves realmente.
Y nada *en el campo visual* permite inferir que es visto por un ojo."
El sujeto metafísico es en este contexto "el sujeto pensante, representante" (5.631), aquel cuya constitución determina lo que puede ser pensado y representado igual que la constitución del ojo determina lo que puede ser visto. Lo que Wittgenstein nos dice es que ese sujeto no puede ser pensado o representado porque no puede aparecer en el mundo que representa y sobre el que piensa al igual que el ojo no puede aparecer en su campo visual (5.6331). El sujeto, como el ojo, sólo determina o conforma los límites de algo, del mundo en un caso (lo pensable o representable) y de la visión en otro (lo visible), pero esos límites *no aparecen* en el mundo o en el campo visual respectivamente, *no pueden ser* pensados, representados o vistos. Wittgenstein va mucho más allá al afirmar en 5.633 que los límites *no pueden ser reconstruidos* a partir de lo que aparece en el mundo o en el campo visual, "nada *en el campo visual* permite inferir que es visto por un ojo". El parágrafo 5.634 supone de hecho un ataque a la posibilidad de un conocimiento a priori, independiente de la experiencia (y, si se quiere, trascendental) de los límites o condiciones de posibilidad de nuestro mundo, que están *fuera* de él. Por eso 5.631 comienza diciendo "El sujeto pensante, representante no existe." Sólo podemos decir que existen las cosas que aparecen en nuestro mundo, en el espacio de lo representable, de lo pensable, todo lo demás no puede pensarse con sentido. Como el sujeto metafísico no aparece en el mundo, no puede ser representado o pensado, no podemos afirmar su existencia. Tal vez pueda describir un cuerpo que veo y que considero *mío* por obedecer a mis pensamientos, pero no puedo describir al "yo que ve" o que "representa" mi cuerpo como mío. Ese yo es un límite *externo* al mundo.

nuestro conocimiento), pero como ése no es el caso, la empresa está destinada al fracaso.

Pese a todo, la imagen metafísica de las paradojas como límites cognitivos parece tener sentido. Tal vez no podamos delimitar cuáles son nuestros límites cognitivos pero, del mismo modo que nuestro fracaso a la hora de resolver una paradoja no demuestra que la paradoja sea irresoluble, nuestra incapacidad para delimitar los límites de nuestro conocimiento no demuestra que nuestro conocimiento carezca de límites. Y si el uso de nuestra capacidad de razonamiento con el fin de obtener conocimiento tiene límites, es plausible que existan paradojas genuinas, entendidas como problemas racionalmente irresolubles, que marquen esos límites, aunque el conocimiento de dichos límites no pueda ser humano o racional. Esta idea nos empuja por un lado a no descartar la posible existencia de paradojas genuinas entendidas como límite y, por otro, a no aceptar la posibilidad de un argumento que identifique los límites de nuestro conocimiento.

Cuando articulamos las dos imágenes metafísicas de las paradojas se plantea un nuevo dilema cognitivo (otro callejón sin salida) que gira en torno al hecho, ya señalado con anterioridad (y al que llamaré HP), de que "no podemos *conocer* la existencia o inexistencia de paradojas genuinas ya que la posibilidad de *demostrar* que una paradoja es irresoluble resulta inconcebible." El dilema es el siguiente: Por un lado, tenemos la obligación de perseverar en la búsqueda de una solución a las paradojas, ya que toda paradoja es sentida como un absurdo inaceptable y, dado HP, aceptar una paradoja supone el riesgo de aceptar falsedades irracionales. Por otro lado, descartar la posibilidad de que algunas paradojas no resueltas sean límites racionales genuinos podría alejarnos de la verdad (en vista de lo que establece HP) obligándonos a perseguir una falsedad: la idea de que nuestra razón puede dar cuenta de ciertos problemas (cuando en realidad sus conclusiones al respecto son erróneas y generan a lo sumo ilusiones metafísicas).

Este nuevo dilema (de "tercer" orden podríamos decir) no nace directamente del problema fundamental y *a priori* que suponen las paradojas ni del examen *a posteriori* de nuestros intentos fallidos por solucionar dicho problema. Este nuevo dilema surge cuando añadimos a los anteriores problemas uno nuevo, la incapacidad de generar una imagen metafísica coherente de la significación que tienen las paradojas en nuestra visión del mundo y del conocimiento humano. *Hasta la fecha*, los intentos de solución de ciertas paradojas han generado paradojas y los intentos de explicar por qué las soluciones a las paradojas generan paradojas (así como el significado que este hecho tiene para nosotros), *también* han generado paradojas. No es de extrañar que, frente a este panorama desolador, descendamos unos cuantos peldaños más la escalera de la justificación racional del conocimiento. Las paradojas plantean problemas concretos a diversas disciplinas científicas o cognitivas. Nuestro intento por solucionarlas dentro de esas disciplinas no es satisfactorio, por eso intentamos comprender o racionalizar de alguna manera el significado que tienen estos fracasos *fuera de ellas*. Por desgracia, nuestros esfuerzos por introducir un poco de orden en este caos generan contradicciones y paradojas metafísicas que hacen incoherente nuestra explicación. La última solución que nuestra *razón* ofrece al problema (cuando se harta de seguir cauces que ya no juzga seguros) adquiere un carácter "religioso" y casi "místico".

Tanto hoy en día como en el pasado ciertas paradojas y problemas intelectuales cuya solución resulta urgente pero inaccesible para aquellos que se han enfrentado a ellos han sido *sentidos* como problemas irresolubles por la razón humana. En la medida en que necesitamos certezas para actuar incluso cuando no disponemos de conocimiento seguro estamos dispuestos a creer ciertas "verdades" de forma irracional. Lo irracional no es tanto el hecho de que formemos creencias en esas circunstancias –después de todo estamos condenados a actuar incluso en la ignorancia–, sino el modo que tenemos de *justificar "la verdad"* de esas creencias.[10] La filosofía no ha escapado tampoco a las tentaciones "místicas" y no es preciso remontarse a los pitagóricos y a su visión de los números irracionales para cerciorarse de ello. La reacción de muchos autores frente a la persistencia de ciertas paradojas y a su reproducción en ámbitos cada vez más generales adopta a veces un carácter radical. No se trata ya de intentar justificar "la verdad" de ciertas contradicciones de forma "coherente", esto es, cuestionando leyes lógicas como la de "*ex contradictione quodlibet*" o justificando que la falsedad de una contradicción no elimina su verdad ni por tanto la validez del razonamiento que la infiere. Se trata directamente de aceptar como inevitable la incoherencia de aquellos discursos que versan sobre paradojas genuinas o límites cognitivos, incluido (como no) el propio discurso que *revela de algún modo* ese hecho.

Llegados a este punto varias son las opciones que podríamos tomar (seguramente no incompatibles entre sí y quizá íntimamente relacionadas). Una primera opción sería aceptar que la razón demuestra contradicciones y que no puede consiguientemente sustraerse a los efectos que racionalmente podemos inferir a partir de una contradicción, a saber, *la verdad de todo* (el trivialismo). Esto implicaría, en particular, la verdad de la tesis que *niega* que todo sea verdadero (la negación del trivialismo). Otra opción sería no otorgar sentido a aquellos discursos en los que la inferencia de contradicciones es inevitable. En ese caso, podríamos decidir guardar silencio al respecto (no hablar sobre lo que carece de sentido). Ahora bien, tan pronto como intentemos *señalar este hecho*, romperemos nuestro silencio para decir cosas sin sentido. En cierto modo, ambas posturas son caras de una misma moneda, se puede llegar al "mutismo" cuando uno cobra conciencia del "sinsentido" del trivialismo y se puede llegar al trivialismo cuando uno intenta justificar el mutismo. El prólogo y el parágrafo final del Tractatus ("De lo que no se puede hablar hay que callar") sugieren actitudes "místicas" de este tipo. Wittgenstein parece optar por el mutismo frente a las irresolubles paradojas que genera el discurso sobre la forma lógica (i.e., aquello que hace posible el significado al ensamblar mundo, pensamiento y lenguaje). Como discurso sobre la forma lógica, el Tractatus *mostraría* a través de

[10] Dar una explicación *racional* es dar una explicación que pueda ser aceptada por todos los seres con uso de *razón* mediante el uso de ésta. Eso es precisamente lo que *no* consigue la justificación de estas creencias.

sus propias contradicciones que todo discurso sobre este tema carece de sentido y coherencia y ha de ser, pues, abandonado una vez cobramos conciencia de este hecho. Nada podemos *decir* con sentido sobre la forma lógica.

En el presente trabajo no abordaremos el problema de la significación que tienen las paradojas en nuestra visión del mundo, i.e., lo que hemos llamado el problema metafísico (ni tampoco el "religioso"). La segunda parte de este trabajo estará dedicada a dilucidar en qué sentido las principales soluciones de que disponemos a paradojas concretas, ya partan de posiciones *doxásticas* (capítulo 4) o de posiciones *ónticas* (capítulo 5) no son satisfactorias. No es mi intención resolver las paradojas, ni siquiera desacreditar una de estas visiones frente a la otra, sino más bien reflejar cómo nuestras reflexiones sobre ciertos temas basculan entre ambas tendencias (a veces de forma aparentemente interminable). Sin embargo, la primera parte de este trabajo (capítulos 2 y 3) estará dedicada a otro problema que aún no hemos descrito y que quizá resulte menos estéril que el de la solución efectiva de las paradojas.

¿Es posible describir las paradojas de un modo uniforme? Esta pregunta encuentra una respuesta obvia en todo diccionario, *cualquier* paradoja puede ser descrita mediante las observaciones hechas en la primera sección de este capítulo. Pero aquí no buscamos *esta* respuesta, i.e., una descripción general del problema que suponen las paradojas: "La existencia de un razonamiento *válido* que *justifica* enunciados falsos". Lo que buscamos aquí es responder a otra pregunta: "¿Tienen los *razonamientos* paradójicos una estructura común?" Deseamos encontrar un patrón estructural que nos ayude a describir *cómo se produce* una paradoja, *qué forma* tiene un razonamiento que toma por conclusión una contradicción. De este modo podríamos avanzar en nuestra comprensión del problema y, lo que aún es más básico, podríamos *identificar* con precisión *un problema común*, podríamos mostrar que, bajo la diversidad aparente de los razonamientos que justifican conclusiones falsas, existe una unidad esencial de fondo. Ciertos elementos se repiten constantemente y se relacionan de un modo similar para establecer una conclusión inaceptable.

Sería francamente sorprendente que todas las paradojas tuviesen una estructura común. Nosotros no defenderemos esta tesis (aunque tampoco aportaremos argumentos de peso en su contra). En nuestro trabajo nos centraremos en una serie de paradojas que los especialistas han denominado paradojas de autorreferencia o reflexividad (siguiendo a Russell 1908) y cuya estructura ha sido descrita recientemente por Priest (2002) basándose en trabajos previos de Russell (1905b). Priest proporciona un esquema en el que tiene cabida un amplio número de paradojas de este tipo (aunque alguna presenta problemas). La función esencial de este esquema será proporcionar un análisis estructural que resulte iluminador con respecto al problema común que encierran estas paradojas.[11] Las paradojas de autorreferencia son particularmente

[11] La ausencia más significativa en este trabajo será la de las paradojas de sorites. Ignoraremos estas paradojas, no porque estemos convencidos de que no puedan encajar en el esquema de Priest, sino porque en su caso, la posibilidad de que encajen o no no es quizá tan importante a la hora de explicar el problema de fondo que suponen como nuestra capacidad de explicar el uso que en ellas hacemos del principio de inducción aritmética.

relevantes para nuestro estudio porque parecen tener la misma estructura que ciertos problemas filosóficos de carácter más general pero igualmente aporético. La idea fundamental es la siguiente: Toda paradoja reflexiva encierra un problema de circularidad viciosa y un problema de diagonalización (ambos rasgos serán descritos en los capítulos 2 y 3). Todas ellas involucran una totalidad absoluta Ω que reúne al conjunto de las cosas que satisfacen cierta propiedad, φ (ser verdaderas, cognoscibles, pensables, etc.), y que podemos generar mediante la iteración de ciertas operaciones (razonamientos, investigaciones, pensamientos o simplemente algoritmos de cierto tipo). Una de esas operaciones, δ, define un elemento de Ω, $\delta(\Omega)$, a partir de *todos* los elementos de Ω, de tal modo que no podemos concebir Ω sin considerar $\delta(\Omega)$ como uno de sus elementos, pero tampoco podemos construir $\delta(\Omega)$ sin definir previamente *todos* los elementos de Ω (incluido $\delta(\Omega)$). Se genera así un círculo vicioso que se agrava cuando nos percatamos de que, debido a la naturaleza de δ, $\delta(\Omega)$ debe *ser* un miembro de Ω y *no serlo*.

Aunque la significación de las paradojas en la filosofía o en nuestra visión del mundo no sea el objeto de estudio de este trabajo,[12] sí es la razón principal que tenemos para abordar el tema de las paradojas y para intentar solucionar problemas que, por sí mismos, no parecen sino acertijos divertidos o incluso ridículos. Lo que está en juego es la posibilidad de justificar absurdos racionalmente y, dada la existencia de principios aparentemente válidos que establecen el trivialismo a partir de

[12] Para un estudio realmente sugerente, amplio e iluminador de este tema consúltese Priest 2002.

la inferencia de algo absurdo, la posibilidad de justificar el uso de la razón en la obtención de conocimiento. Quizá el problema que suponen las paradojas para la lógica no sea más grave que el problema que supone el escepticismo para la epistemología pero, en cualquier caso, *no es menos grave* (de hecho, el razonamiento escéptico *es paradójico*). Tal vez podamos ignorar los argumentos escépticos y las paradojas en nuestra vida cotidiana, pero no cuando intentamos comprender qué significa conocer o qué significa razonar. Por este motivo, dedicaremos este capítulo a ilustrar mediante un ejemplo lo que hemos intentado exponer de forma más abstracta, general y vaga en estas páginas iniciales: cómo las paradojas generan problemas filosóficos e imágenes del mundo relacionadas con nuestra comprensión de las capacidades y necesidades cognitivas humanas.

1.3. Dos imágenes de las paradojas (I): paradojas y progreso cognitivo

Las paradojas han reclamado la atención de la filosofía desde el principio, no en vano la palabra que utilizamos hoy en día para designarlas es de origen griego. Los ejemplos son abundantes: los razonamientos de Zenón sobre la imposibilidad del movimiento, la paradoja del mentiroso, la del tercer hombre, las paradojas de sorites y un largo etcétera. Muchas de estas paradojas han llegado intactas a nuestros días y otras son ancestros aún reconocibles de paradojas de "nuevo cuño". Cuando miramos hacia atrás en la historia nos damos cuenta de que muchas paradojas han evolucionado hacia formulaciones más precisas e inquietantes y otras han desaparecido como problemas inherentes a esquemas conceptuales inadecuados que hemos superado con el tiempo. Frente a un razonamiento inaceptable pero aparentemente válido que establece una conclusión falsa a partir de premisas verdaderas, cabe una única reacción inicial en vista de las consecuencias discutidas al principio de 1.2. Tenemos la "obligación" de mostrar que el argumento es falaz o que la conclusión es en el fondo verdadera.

En el primer caso, hemos de probar la falsedad de las premisas o la incorrección de las reglas de inferencia. En el segundo, hemos de aceptar la verdad de la conclusión que nos impone el argumento. La paradoja es verdadera en contra de lo que parece y es nuestra opinión común o más inmediata acerca de lo que enuncia la que se halla equivocada. Un ejemplo de la primera estrategia sería la solución que la teoría de conjuntos impone a la paradoja de Russell [2]. La paradoja se entiende como una *reductio ad absurdum* de todos aquellos supuestos y principios que permitan la construcción de un conjunto que tenga por elementos exactamente a aquellos conjuntos que no se pertenecen a sí mismos[13]. Un ejemplo de la segunda posibilidad es el teorema de Gödel del que se sigue el paradójico resultado de que "no todas las verdades de

[13] El "Esquema axiomático de separación" es, dentro de las axiomatizaciones clásicas de la teoría de conjuntos, el principal filtro para la paradoja de Russell. El esquema nos permite reunir en un conjunto x a todos aquellos conjuntos y que satisfacen una propiedad dada, P. Sin embargo, para evitar que se construya el conjunto R de todos aquellos conjuntos y que satisfacen la propiedad de "no pertenecerse a sí mismos", el esquema nos dice que los y agrupados en el conjunto x, deben ser miembros de un conjunto *previamente existente* o *construido* z. Esto es, x ha de ser el conjunto de todos los y, *pertenecientes a* z, que satisfacen la propiedad P. Este problema se discutirá en el capítulo 4 (4.4).

la aritmética son demostrables".[14] En este caso nuestras intuiciones de partida nos dicen que el enunciado es falso, sin embargo el teorema de Gödel se interpreta como un razonamiento válido que establece la verdad del enunciado mencionado, lo que nos fuerza a revisar cualquier argumento o justificación racional que pudiésemos tener a favor de la falsedad de dicho enunciado.

Desde esta perspectiva, solucionar una paradoja consiste, pues, en mostrar que en el fondo la paradoja no existe o no es genuina: o bien existe un error en nuestro argumento (inferencias o premisas), o bien existe un error en nuestras creencias sobre la conclusión. En el peor de los casos el error es inherente a la teoría que avala nuestro argumento pero entonces la teoría se rechaza, lo que implícitamente supone que la paradoja "no puede ocurrir" en una teoría correcta. No hay teorías correctas que justifiquen la posibilidad de un argumento válido con premisas verdaderas y conclusión falsa. Es este continuo proceso de revisión de nuestras teorías valorando distintas posibilidades en busca de

[14] En dicho teorema se construye un sistema de codificación (i.e., gödelización) que permite expresar oraciones de la meta-aritmética mediante oraciones de la aritmética. El teorema identifica una oración, A, de la aritmética cuyo significado es reinterpretado en términos de la oración meta-aritmética: "A no es demostrable". Vemos que si A es demostrable, entonces *es falsa* (lo que resulta absurdo), por tanto A no puede ser demostrable. Ahora bien, tampoco su negación, $\neg A$, es demostrable, ya que, de serlo, *sería verdad* (por la ley de doble negación: $\neg\neg\alpha \leftrightarrow \alpha$) que A es demostrable, y hemos visto que eso es imposible. Así pues, ni A ni $\neg A$ son demostrables y la aritmética es, consiguientemente, incompleta. En la medida en que A no es demostrable, A es además *verdadera* porque eso es precisamente lo que enuncia. Por lo tanto, *algunas verdades de la aritmética no son demostrables*. Discutiremos la relación que guarda el teorema de Gödel con las paradojas en el capítulo 5 (5.2.1.1 y 5.3.7).

una solución satisfactoria a las paradojas lo que explica el progreso científico.[15] Esta imagen se puede ilustrar mejor quizá con un ejemplo.

Pensemos en el famoso poema hexamétrico de Parménides.[16] En él, empujado por "el derecho y la justicia" (fr. 1, verso 28), Parménides cuenta como se aparta del "trillado sendero de los hombres" (fr. 1, v. 27) para seguir un camino por el que será conducido en pos de la "Verdad" más allá de un mundo cuyos límites imponen las puertas de "la Noche y del Día" (fr. 1, v. 11). Al otro lado de las puertas, que presumiblemente delimitan el mundo sobre el que *opinan* los hombres, una diosa lo recibe y le revela las dos "únicas vías de investigación pensables" entre las que

---

[15] Un magnífico compendio de esta visión de las paradojas se puede encontrar en Quine 1976b. Allí, Quine describe como paradojas verídicas o falsarias a aquellas cuya solución –i.e., la conversión de su argumento principal en una *reductio ad absurdum* de algún supuesto de partida– llega a imponerse con naturalidad tras la sorpresa inicial que nos causa. (En una paradoja verídica reconocemos la verdad de lo que creíamos falso inicialmente; en una paradoja falsaria el proceso se invierte y acabamos aceptando la falsedad de lo que nos parecía verdadero.) Como ejemplos de paradojas "domesticadas" Quine considera la solución a las paradojas de Zenón, el teorema de Cantor, los teoremas de incompletud de Gödel y lo que hoy nos parece un rasgo esencial de toda colección infinita, A: la existencia de una correlación entre A y alguno de sus subconjuntos *propios*. Así mismo, Quine reserva el nombre de "antinomia" a lo que aquí hemos denominado a veces "paradojas genuinas", i.e., aquellas cuya conversión en una *reductio* no nos satisface por parecernos arbitraria y *ad hoc*. Quine Tampoco olvida la dimensión práctica del problema al mencionar (y suscribir) la distinción en familias de Ramsey (véase 2.2). Por último, fiel a las convicciones defendidas en textos como "Two dogmas" (1980b), Quine no descarta que lo que hoy nos parece una antinomia (por poner en jaque algunos de los principios más hondamente arraigados en nuestro esquema conceptual) pueda parecernos en el futuro, tras un cambio de esquemas conceptuales forzado por nuestra reflexión sobre estos problemas, una paradoja verídica o falsaria. Esto es lo que ocurrió, según Quine, con las paradojas de Zenón. La revisabilidad de nuestras convicciones (también sobre las paradojas) es un hecho histórico.

[16] Cito los fragmentos (fr.) seguidos de los versos (v.) según la numeración de H. Diels: *Die Fragmente der Vorsokratiker*, 5ª edición, ed. W. Kranz, Berlin, 1934-1954. Así mismo, sigo en algunos de mis comentarios la interpretación que hacen de estos

habrá de escoger: "La una, que es y que le es imposible no ser, es el camino de la persuasión (porque acompaña a la Verdad); la otra que no es y que le es necesario no ser [...] es una vía totalmente indiscernible; pues no podrías conocer lo no-ente (es imposible) ni expresarlo." (fr. 2 v. 2-8).[17] Dada la naturaleza de ambas vías y dado que "[l]o que puede decirse y pensarse debe ser, pues es ser, pero la nada no es" (fr. 6, v. 1-2), la elección parece clara. Una de las dos vías es impensable *porque no es*[18] (ni puede ser), así pues, si la otra es pensable, *debe ser*, y es la única además que podemos elegir habida cuenta de que ninguna de las vías se puede reducir a la otra: "nunca se probará que los no entes sean" (fr. 7, v. 1). Así mismo, debemos rechazar todas aquellas opiniones que tengan como consecuencia la afirmación de que *lo que es, no es*; o la afirmación de que *lo que no es, es*. Debemos evitar ser iguales a las "gentes sin juicio, que creen que ser y no ser son lo mismo y no lo mismo" (fr. 6, v. 7-9). Proscribiendo toda contradicción a partir de estos principios, Parménides deduce la naturaleza de "lo que es", apuntando que "la decisión en estas cosas se basa en esto: es o no es." (fr. 8, v. 15-16). Con respecto a "lo que es" nos dice que "por ser ingénito, es también

---

fragmentos Kirk, Raven y Schofield 1983, a cuya versión española pertenecen las citas traducidas y el resto de referencias.

[17] Algunos editores completan el verso 8 con: "Pues lo mismo es ser pensado y ser." (fr. 3). Sin embargo, véase Kirk, Raven y Schofield, op. cit., p. 357 (nota 2).

[18] La idea parece ser, como señalan Kirk, Raven y Schofield (op. cit., p. 355), que para pensar en algo o expresar algo ese "algo" debe ser o existir de algún modo, en caso contrario no habría nada *en* lo que pensar ni, alternativamente, nada *sobre* lo que expresar un pensamiento completo. Si uno de los términos referenciales de una oración es "vacío" (no tiene referente) ésta no expresa una proposición (una idea muy arraigada en la tradición analítica desde Russell y Frege, quien, para evitar problemas de este tipo, asignó a todo término "vacío" un referente arbitrario: el "conjunto vacío").

imperecedero, entero, monogénito, inmóvil y perfecto" (fr. 8, v. 3-4). Cualquier otra posibilidad iría en contra de los principios establecidos por Parménides. "Lo que es" no puede, por ejemplo, cambiar, llegar a ser o dejar de ser porque para ello tendría que contener a (o venir de, o convertirse en) "lo no-ente", que como hemos visto *no es*, por lo tanto "lo que es" ni cambia, ni tiene comienzo ni fin (está fuera del tiempo). Argumentos similares se aducen en contra de la divisibilidad, la discontinuidad o la imperfección de "lo que es".[19]

Tenemos así todos los ingredientes de una paradoja: 1) Un razonamiento aparentemente válido que, aplicando el principio de no contradicción a ciertos objetos de discurso ("lo que es" y "lo que no es"), establece 2) una conclusión que choca con nuestras convicciones y que percibimos como extraña o falsa: No son posibles ni la pluralidad ni el cambio ni el movimiento; no hay muerte ni nacimiento; nada es divisible ni imperfecto. Tenemos también 3) una reacción, un intento de solución al conflicto que genera la paradoja.

Consciente de lo inusitado de sus conclusiones pero convencido también de su verdad, Parménides ataca nuestras convicciones empíricas (la fuente de nuestra extrañeza) y descalifica "las opiniones de los mortales" en las que no hay "creencia verdadera" (fr. 1, v. 30). Para ello, reviste de autoridad sus ideas apelando a la diosa por cuya boca habla y a la legitimidad del camino por el que le empujan el "derecho" y la "justicia", (donde ésta es presentada a la vez como el acicate de las

[19] Todos los argumentos se concentran en el fragmento 8 del poema según la edición de Diels.

reflexiones de Parménides y lo que mantiene firme y encadenado a su naturaleza a "lo que es" -fr. 8, v. 14-15).[20] Nuestras creencias son presentadas en cambio como fruto del "hábito, hijo de la mucha experiencia" y no debemos permitir que éste nos arrastre hacia la vía que "no es" (fr. 7). Parménides nos urge, pues, a aceptar la verdad de sus razonamientos frente a la falsedad última de nuestros prejuicios basados en la experiencia. Ésa es su solución al conflicto que plantea la paradoja.

Alternativamente, nos encontramos con otra solución histórica, de signo opuesto, según la cual lo erróneo es el argumento de Parménides ya que asume falsamente que "lo *que es* y lo *uno* significan lo mismo, mientras que otros deshacen el argumento de Zenón y Parménides al afirmar que lo *que es* y lo *uno* se dicen de muchas maneras".[21] En la *Metafísica*, Aristóteles, que se encuentra entre los detractores de Parménides, insiste a menudo en que "'lo que es', sin más precisiones, se dice en muchos sentidos: en primer lugar, está lo que *es* accidentalmente; en segundo lugar, lo que *es* en el sentido de <<es verdadero>> y lo que *no es* en el sentido de <<es falso>>; además, están las figuras de la predicación" –lo que Aristóteles llamará *categorías*– "(por ejemplo qué [es], de qué cualidad, de qué cantidad, dónde, cuándo [es], y cualquier otra cosa que signifique de ese modo, y aún, además de todos estos [sentidos], lo que *es* en potencia y en acto" (*Metafísica* VI, capítulo 2, 1026a,33 - 1026b,1). Todas estas maneras de decir y entender lo que *es* revelan que el problema de Parménides consiste en haber *identificado*

[20] Si bien es cierto que en el primer caso se habla de la "justicia" y en el segundo de la "Justicia".

erróneamente cosas *homónimas*[22]. Por eso Aristóteles puede salvar el principio de no contradicción como el más básico de todos (*Metafísica* IV, capítulo 3) y negar, no obstante, las conclusiones de Parménides sobre la contradictoriedad de la pluralidad de entes, del cambio, del movimiento, etc.: La semilla, que *no es* un árbol *en acto*, *es* un árbol *en potencia*; y no hay nada extraño en que algo que *era* un ser vivo *ya no lo sea*; o en la verdad de que haya cosas que *sean* árboles y cosas que *sean* peces, lo que no impide que los árboles *no sean* peces o que algo *sea* un árbol y *no sea* alto. Sólo "*es imposible que lo mismo se dé y no se dé en lo mismo a la vez y en el mismo sentido* (y cuantas precisiones habríamos de añadir, dense por añadidas frente a las dificultades dialécticas)." (*Metafísica*, IV, capítulo 3, 1005b, 18-21). Lo que no puede ocurrir, por ejemplo, es que *lo mismo sea y no sea* un árbol *en acto*, o verde y no verde *a la vez*; o que algo que *es esencialmente* un hombre, *no lo sea y sea esencialmente* un perro. El problema sólo aparece cuando utilizamos contradictoriamente el término 'ser' aplicado a lo mismo y en el mismo sentido y circunstancias.

Independientemente del éxito o, a su vez, de los problemas que puedan nacer de esta solución, lo que llama la atención en la respuesta aristotélica a la paradoja es la riqueza conceptual del análisis con que

[21] Aristóteles, *Sobre las refutaciones sofísticas*, 182b, 25-28.

[22] "Se llaman *homónimas* las cosas cuyo nombre es lo único que tienen en común, mientras que el correspondiente enunciado de la entidad es distinto, v.g. *vivo* dicho del hombre y dicho del retrato" (*Categorías*, 1a). Es cierto que existe una fuerte conexión entre los distintos sentidos en que se entiende *lo que es*: "La expresión 'algo que es' se dice en muchos sentidos, pero en relación con una sola cosa y una sola naturaleza y no por mera homonimia" (*Metafísica*, IV, capítulo 2, 1002b, 33-5). Pero, pese a todo, esa

sustituye el modo comparativamente tan simple y directo que tiene Parménides de abordar el hecho de que algo "sea". Precisamente porque las paradojas son inaceptables, a menudo tienen este efecto reestructurador en nuestras creencias. Las paradojas muestran que los conceptos, prácticas y convicciones que gobiernan ciertos ámbitos de nuestra vida pueden ser incoherentes. El desasosiego que esto produce sólo se puede paliar rechazando nuestras convicciones sobre la falsedad de ciertos enunciados, (como nos pide, en este caso inverosímilmente, Parménides); o cambiando las premisas, o el marco conceptual (como nos pide Aristóteles), o la estructura inferencial (la lógica) de los que se derivó una falsedad. El teorema de Gödel y la paradoja de Russell planteaban escenarios similares.

De estas observaciones se desprende una imagen positiva y bastante extendida de la función que las paradojas han desempeñado tradicionalmente en el marco de las ciencias y de cualquier otra disciplina teórica (incluida la filosofía). Las paradojas favorecen el progreso de la ciencia y del conocimiento en general en la medida en que constatan incoherencias en nuestras convicciones sobre el mundo y nos fuerzan a revisar nuestros juicios sobre ciertas materias o incluso a reestructurar los principios lógicos, teorías y esquemas conceptuales bajo los que concebimos dichas materias y razonamos sobre ellas. El resultado de estos reajustes ha de ser una mejora, un desarrollo de nuestros instrumentos conceptuales para representar el mundo de forma coherente

---

conexión no hace *idénticos* a los distintos sentidos en que podemos decir de algo que "*es*".

y más precisa, sin paradojas. Quizá los reajustes sean conservadores y tiendan siempre a sacrificar el menor número posible de convicciones previas, eligiendo preferentemente aquellas que ocupen una posición menos central en el contexto de nuestras creencias. Pero, en cualquier caso, los reajustes tendrán lugar y serán tan drásticos como sea necesario para evitar que un argumento válido no preserve la verdad de las premisas en la conclusión; aunque para ello tengamos que calificar de "aparentes" las verdades que fundamos en nuestra experiencia sensorial (como hace Parménides) o tengamos que reconstruir mediante un análisis detallado nuestra comprensión de un concepto problemático enriqueciéndolo en sentidos y distinciones (como Aristóteles).

Ésta es una imagen de las paradojas sin duda poderosa. Lo que esta imagen nos dice es que cuando la ciencia "desenreda" con éxito una paradoja, nuestro conocimiento y nuestra capacidad de representar el mundo con coherencia y precisión mejoran, pero eso sólo es posible si la paradoja *desaparece*. En la medida en que tiene sentido construir un discurso científico alrededor de una paradoja, ésta debe tener también *solución* ya que no cabe decir nada coherente sobre campos en los que las contradicciones son irresolubles. Ninguna teoría razonable podría aceptar la verdad de una contradicción. No puede haber ciencia donde hay círculos cuadrados, donde es imposible para el ser humano alcanzar soluciones o respuestas a sus preguntas. Las paradojas sólo tienen cabida en el discurso científico en la medida en que son percibidas como acertijos, ilusiones o rompecabezas por resolver, si esta percepción se desvanece, entonces la paradoja pasa a ser contradicción y el discurso al que aparece inexorablemente ligada cae bajo sospecha. Al menos formalmente, ésta es la acusación que pesa sobre muchos ámbitos teóricos: 1) No puede existir acuerdo objetivo allá donde las contradicciones son irresolubles; 2) Las contradicciones no pueden pertenecer "al mundo" (no hay hechos contradictorios), sino sólo a nuestros juicios y maneras de representar "el mundo"; 3) Por tanto, allá donde el acuerdo general sobre los hechos es inalcanzable debido a la presencia *necesaria* de contradicciones, no existe objetividad, no existen hechos sobre los que construir un acuerdo ni por ende ciencia o conocimiento. Consiguientemente, todo discurso que acepte una contradicción es expulsado del ámbito de la ciencia. En gran medida, la lucha por reivindicar el valor de ciertas disciplinas teóricas ha consistido en buscar en su seno áreas de posible consenso sobre las que fundar un marco de objetividad; pero también en situar toda contradicción fuera de su ámbito de discurso y en reinterpretar aquellas que parecen inevitables como algo más que simples contradicciones o incoherencias metodológicas.

1.4. Dos imágenes de las paradojas (II): paradojas y límites cognitivos

Volvamos de nuevo al poema hexamétrico. Hay en los fragmentos de dicho poema otra paradoja que Aristóteles no parece tematizar y en la que ni siquiera el propio Parménides repara.[23] Cuando Parménides se

[23] Como indicamos en la nota 18, Kirk, Raven y Schofield sí se percatan de ella y esbozan el problema que aquí discutimos más extensamente.

pregunta en general sobre lo que se puede conocer, no sólo llega a la conclusión de que sólo se puede conocer "lo que es", sino que juzga además que *sólo se puede pensar y expresar* "*lo que es*".[24] Incluso para alguien no comprometido con las tesis de Parménides, esta afirmación (o una similar) parece factible. Si podemos pensar *en* algo, entonces debe haber algún sentido en el que ese algo *"sea"*, debe haber algún sentido en el que ese algo exista, aunque sólo sea como objeto o contenido de nuestros pensamientos (quizá como simple alucinación, o quizá como invención, o imagen mental: la de un caballo alado llamado 'Pegaso', por ejemplo). A veces nos equivocamos con respecto a cuál es el objeto o contenido de nuestros pensamientos o con respecto a su naturaleza, a veces tenemos pensamientos falsos, pero lo que no parece plausible es que podamos tener pensamientos *sobre* "lo que no es" en ningún posible sentido (o en el único, según Parménides) del término 'es'. ¿Hacia *qué* dirigiríamos nuestros pensamientos entonces? Quizá creamos erróneamente que Don Quijote o Hamlet son personas reales pero parece claro en todo caso que, si podemos hablar *de ellos* o pensar *en ellos* (aunque sea erróneamente), es porque *son* algo. Por eso decimos cosas tales como: "Creía que Don Quijote era un hombre de carne y hueso pero *es* sólo *un personaje de ficción*". Si Don Quijote no fuese nada en absoluto, 'Don Quijote' sería un nombre vacío, sin contenido, y no

[24] Diels: fragmento 2: 7-8; (fr. 3?, véase nota 17); fr. 6: 1; y fr. 8: 8-9, 16-17 del poema hexamétrico de Parménides.

podríamos pensar *en Don Quijote* o hablar *de él*. Si podemos hablar *de él* o pensar *en él* es preciso que *sea algo*.[25]

Estas observaciones hacen plausible la afirmación de que lo que no es, lo que no existe de ningún modo, "no es decible ni pensable" (fr. 8, v. 8, 9). No es pensable porque no hay *nada en que* pensar y no es decible porque, si no es, entonces no hay *nada que* significar o simbolizar mediante el lenguaje. Quien siga el camino de "lo que no es" se verá forzado a "usar una mirada vacilante o un oído y una lengua plenos de sonido sin sentido" (fr. 7, v. 3-5) porque a eso se reduce todo discurso sobre lo que no es, a sonidos sin sentido.

Pero, si esto es así, *¿a qué* se refiere Parménides cuando habla de "lo no-ente" o "lo que no es" a lo largo del poema o cuando dice en el fragmento 2 que sólo hay dos vías de investigación "*pensables*": la que es

---

[25] El problema de cómo se puede hablar sobre "lo que no es" o "no existe" y del significado que puedan tener oraciones con términos cuya referencia es dudosa todavía nos acompaña hoy en día y no es, ni mucho menos, un problema trivial, ya zanjado. Todo lo que nos interesa decir aquí acerca de esto es 1) que el poema de Parménides apunta, en parte, este problema; y 2) que, al igual que él, hoy en día (y en el pasado) los filósofos han juzgado inaceptable la referencia a entidades que "no son" en ningún posible sentido. Uno de los ejemplos históricos más famosos es el enfrentamiento de Russell con Meinong en "On Denoting". Según el segundo, entidades como "el círculo cuadrado" deben "subsistir" o "ser" en algún sentido, en caso contrario la expresión 'el círculo cuadrado' no sería significativa y oraciones como 'El círculo cuadrado no existe' (trivialmente verdaderas) carecerían de significado por contener un término denotativo vacío. La intuición de Russell, en cambio, es que no hay ningún sentido concebible en el que los círculos cuadrados puedan *ser*, *existir*, *subsistir* o ser referidos. Su solución a la significatividad de la oración anterior pasa por analizar expresiones denotativas (descripciones definidas, por ejemplo) de tal modo que no presupongan la existencia de un referente y sean, pese a todo, significativas. Lo relevante de la discusión es: 1) que ambos autores están de acuerdo en que las expresiones referenciales *sin referente* no tienen significado; y 2) que "lo que no existe" o "no es" en ningún sentido posible no puede ser referido. Sus diferencias afectan sólo a sus análisis en el terreno de la ontología y la semántica.

y *la que no es*? ¿*En qué* hemos de pensar, *qué* designan sus palabras? Parménides parece hablarnos de aquello de lo que nos ha prohibido hablar, intenta referirse a "lo no-ente" para decirnos precisamente que "no es". Pero, ¿cómo decir algo con sentido acerca de "lo no-ente", incluso algo tan básico como que "no es"? El discurso sobre "lo no-ente" es aporético. Al identificar "lo no-ente" con lo que queda fuera del alcance de nuestro pensamiento y de los límites de cualquier ontología aceptable, Parménides se sitúa sin querer más allá de los límites que él mismo ha descrito. Esa supuesta "identificación" no es posible porque apunta precisamente hacia lo que no tiene condiciones de identidad, hacia lo que, por no existir, no puede ser pensado, referido o identificado.

Tal vez la única manera de aceptar la significatividad de expresiones como 'lo no-ente' o 'lo que no es' en oraciones que las contengan sea interpretar dichas expresiones de tal modo que *no refieran* o *nombren* nada, so pena de contradicción. Podríamos seguir quizá una estrategia de corte russelliano y parafrasear ésta y otras expresiones similares mediante: 'Para todo x, si x no es, entonces x ...'; o mediante: 'Existe un x tal que x no es y x ...'. De esta manera evitaríamos nombres problemáticos poniendo en su lugar expresiones significativas cuyas variables no serían satisfechas por ningún ente, capturando así parte del sentido de la expresión original.[26]

[26] Pero no está en absoluto claro que la solución funcione o sea intuitiva. 'Lo no-ente no es' podría parafrasearse quizá como 'Para todo x, si x no es, entonces x no es'. Los condicionales de la oración serían verdaderos, no por la identidad entre antecedentes y consecuentes, sino por la falsedad de los antecedentes (ya que 'x' sólo es sustituible por cosas que *son*). ¿Y qué ocurriría con 'Lo no-ente es', debería interpretarse como 'Existe un x tal que x no es y x es'? ¿Debería 'Para todo x, si x no es, x es' ser falsa o verdadera?

Una vez más, si esta estrategia es aceptable (cosa discutible), nos encontramos frente a la "solución" de una paradoja. Sin embargo, este caso parece sustancialmente diferente al del ejemplo anterior. Si Aristóteles está en lo cierto, el carácter paradójico del movimiento, la pluralidad de seres y el cambio obedecen a un error de Parménides. En el presente caso, en cambio, aunque hayamos conseguido hablar con sentido acerca de los límites del lenguaje, del pensamiento y de la ontología evitando una paradoja, hemos descubierto que determinados proyectos son necesariamente problemáticos y tal vez imposibles. Quizá podamos hablar con sentido de los límites del lenguaje, pero parece imposible referir "lo que no es" sin transgredir esos límites. Por eso cuando lo intentamos entramos de lleno en el ámbito de lo paradójico. Parménides no se equivoca en esto. Hemos descubierto, además, que al hablar sobre este asunto (incluso para decir cosas que juzgamos ciertas) debemos andar con mucho cuidado porque los límites entre lo paradójico y lo no paradójico son tan delgados que a veces cuesta saber si existen o si es posible trazarlos y hablar de ellos sin cruzarlos irremediablemente.[27]

---

[27] El discurso sobre "lo que no es" tiene de hecho la estructura superficial que hemos asociado a las paradojas reflexivas. Consideremos la totalidad absoluta Ω de "las cosas sobre las que podemos hablar", una manera de determinar que algo, x, pertenece a esa totalidad es construir una oración que hable de x, existen obviamente muchas operaciones lingüísticas que nos permiten formar oraciones que hablen de x. Podemos, por ejemplo, identificar x mediante una descripción definida que aporte una condición necesaria y suficiente para ser x (por ejemplo, x es 'lo que tienen en común sólo los números pares', i.e., *ser divisible entre 2*; o x es 'el padre de Atenea', i.e., *Zeus*). A continuación podemos predicar algo del referente de la descripción definida, i.e., de x ('Ser divisible entre 2 es una propiedad que satisface 4'; o 'El padre de Atenea es el señor del Olimpo'). De esta manera determinamos que x pertenece a la totalidad de las cosas sobre las que podemos hablar. Pero, ¿qué ocurre cuando queremos hablar de *lo que no es*? Básicamente, operamos del mismo modo: identificamos aquello de lo que

La reflexión sobre este tipo de proyectos y sobre otros similares unida a los problemas que de ellos se derivan ha generado una imagen de las paradojas que está estrechamente vinculada a la filosofía. Según esta imagen, hay ciertas paradojas que son radicalmente *irresolubles* y lo son precisamente por estar ligadas a proyectos de este tipo, proyectos cuya viabilidad es imposible por presuponer la validez de contradicciones latentes o por exceder de algún modo los límites de la experiencia y de las capacidades humanas. No es raro que la persecución de estos proyectos desemboque habitualmente en procesos circulares o regresos al infinito. El problema no radica aquí en la manera que tenemos de describir dichos proyectos o de teorizar sobre ellos. Ningún cambio en nuestra forma de presentarlos o concebirlos eliminaría por completo las paradojas. El problema sigue ahí por muchas reformas teóricas que emprendamos. Lo que diferencia a estas paradojas y a los proyectos que las originan del tipo de paradojas y proyectos que hacen avanzar a la ciencia y que ésta acepta como retos en su avance es que las paradojas genuinas son sencillamente irresolubles, no son fruto de ilusiones o errores subsanables. Una consecuencia inmediata de este hecho es que no es posible construir una ciencia o una práctica coherentes allá donde estas

---

queremos hablar mediante una descripción definida: 'lo que no es' y predicamos cosas acerca de *lo que no es*: 'Lo que no es, no es', 'Lo que no es no puede ser pensado o dicho'. De este modo hablamos de lo que no es y mostramos que forma parte de Ω. El problema surge cuando nos percatamos, leyendo a Parménides, de que una condición necesaria para hablar significativamente de algo es que *ese algo* "*sea*", en caso contrario nuestras palabras no tendrían significado (no habría *nada* qué significar). Ahora bien, *lo que no es, no es* y, por tanto, la expresión 'lo que no es' carece de sentido. Esto significa que no podemos hablar de lo que no es (i.e., que no pertenece a Ω). Tenemos así una contradicción.

paradojas surgen, por eso los proyectos a los que aparecen indisolublemente ligadas tienen un estatus problemático. Pero, si esto es así, ¿qué interés puede tener que nos detengamos en ellas?

De acuerdo con esta imagen, la relevancia de las paradojas es doble. Por un lado, nos muestran (de la única manera que parece factible o fiable) cuáles son los límites reales del razonamiento. Si ninguna reforma teórica concebible nos permite eliminar las paradojas dentro de un ámbito o práctica determinados, entonces –concluimos– debe ser radicalmente imposible razonar de forma coherente en dicho ámbito o práctica. La idea es que sólo fracasando sistemáticamente en nuestros intentos por resolver un problema, podemos "entrever" que el problema en cuestión *no tiene solución.* La paradoja funciona aquí como "prueba" de la imposibilidad de algo, a saber, la construcción de una ciencia o de una práctica coherente en cierto ámbito. Es análoga en este sentido al papel que juega la inferencia de una contradicción, dentro de una reducción al absurdo, en la refutación de una premisa o de un supuesto.

Pero no es este papel positivo que las paradojas genuinas desempeñan en cierto tipo de razonamientos lo que les confiere un carácter inquietante. Si estas paradojas nos preocupan es precisamente porque "muestran" que es imposible razonar coherentemente en ciertos ámbitos que, sin embargo, nos parecen irrenunciables. Este aspecto marca el segundo sentido en que son relevantes. Según la imagen de la filosofía que deseamos motivar los seres humanos estamos interesados en obtener conocimiento incluso allá donde, en principio, parece imposible obtenerlo. Nuestra razón se esfuerza por dar respuesta a ciertas preguntas que adquieren a veces la misma forma –aquella bajo la cual tiende la razón a ordenar todo aquello sobre lo que se vierte–, pero que se aplican indistintamente a todo tipo de materias, cobrando paulatinamente un carácter cada vez más general y menos restringido. Cuando nuestras preguntas exceden por su generalidad los límites dentro de los cuales podemos razonar siguiendo principios fiables, nuestros razonamientos se vuelven incoherentes. Sin embargo, y eso es lo inquietante, las preguntas siguen planteándose con urgencia y siguen pidiendo respuestas, con lo que no cejaríamos en nuestro empeño por responderlas aun cuando todo esfuerzo fuese a la postre vano y toda respuesta insatisfactoria.

Uno de los rasgos distintivos de la filosofía y, en particular, de una de sus ramas, la metafísica, es el haberse ocupado tradicionalmente de este tipo de preguntas completamente "generales" y no restringidas, preguntas que afectan incluso a hechos y objetos de los que no tenemos experiencia posible. De ahí la especial significación filosófica que adquieren las paradojas cuando se presentan como el resultado de proyectos que pugnan por responder a preguntas de este tipo. La generalidad de estas preguntas y proyectos se ha utilizado a menudo para marcar la distancia que separa la filosofía de las ciencias. A diferencia de las ciencias particulares, se arguye, la filosofía no tiene un objeto restringido de estudio, por eso resulta particularmente difícil definir sus tareas y su campo de acción y por eso se ha sugerido a veces que no existe una disciplina distinta de la filosofía llamada "metafilosofía". En otros ámbitos quizá quepa distinguir entre una ciencia *t* que tenga por objeto la materia *m* y una (meta-)ciencia distinta *t'* cuyo objeto sea la propia ciencia *t* (pero no *m*). Lo que se afirma aquí es que toda "metafilosofía" es en el fondo filosofía.[28] En la medida en que se relaciona esta generalidad con la formación de paradojas irresolubles, las ciencias han querido también marcar distancias con respecto a la filosofía. Si entendemos la filosofía como ciencia, sin restringir debidamente su alcance, y presentamos sus resultados como conocimiento, tal y como pretende la metafísica, entonces tropezaremos irremisiblemente con contradicciones, paradojas y argumentos insatisfactorios (de donde se sigue el descrédito de la metafísica como ciencia). Si, por contra, restringimos prudentemente el alcance de nuestras cuestiones filosóficas, quizá logremos construir una ciencia, pero esa ciencia nunca podrá ser identificada con la filosofía porque habrá renunciado a parte de las preguntas y ámbitos de reflexión que le son esenciales. Así pues, la filosofía (sin restricciones) se enfrenta a este dilema: O se presenta como ciencia y entonces se convierte en metafísica (en el sentido peyorativo del término); o renuncia a presentarse como ciencia y entonces debe explicar cuál es el estatus de sus afirmaciones, ya que no pueden tomarse por conocimiento en sentido estricto.[29] El

[28] Un conocido ejemplo: "Pudiera pensarse: si la filosofía habla del uso de la palabra <<filosofía>>, entonces tiene que haber una filosofía de segundo orden. Pero no es así; sino que el caso se corresponde con el de la ortografía, que también tiene que ver con la palabra <<ortografía>> sin ser entonces de segundo orden." (Wittgenstein, *Investigaciones filosóficas*, §121).

[29] Una vez más, Wittgenstein y Kant son los referentes más claros. El mejor resumen imaginable de estas ideas se halla condensado en los dos párrafos iniciales (citados a continuación) del prólogo a la primera edición de la *Crítica de la razón pura* (Avii-Aviii):

"La razón humana tiene el destino singular, en uno de sus campos de conocimiento, de hallarse acosada por cuestiones que no puede rechazar por ser problema es, como vemos, grave, esta imagen de la filosofía sufre en sus propias carnes la lacra de las paradojas incluso al reconocerlas como inevitables en su ámbito de reflexión. Pese a todo, esta imagen no es en absoluto gratuita, está presente en las obras de numerosos pensadores desde Kant hasta Wittgenstein y ha llegado con fuerza a nuestros días. Nociones como la de "antinomia de la razón pura" en Kant o la distinción entre "decir" y "mostrar" del *Tractatus* tienen mucho que ver con esta imagen. Resultaría interesante explorar la relación que existe entre esta visión de la filosofía y la estructura de ciertos proyectos filosóficos que asociamos a paradojas recalcitrantes. No abordaremos estas cuestiones aquí, pero sí esbozaremos, a grandes rasgos, la "forma" que suele adoptar el problema.

Decíamos al principio que toda paradoja se manifiesta en el curso de un tipo determinado de acción, i.e., de un razonamiento, es bueno mantener presente este vínculo entre paradojas y acción. El éxito en la ejecución o desarrollo de toda acción, proyecto o práctica dirigidas a la

---

planteadas por la misma naturaleza de la razón, pero a las que tampoco puede responder por sobrepasar todas sus facultades.

La perplejidad en la que cae la razón no es debida a culpa suya alguna. Comienza con principios cuyo uso es inevitable en el curso de la experiencia, uso que se halla, a la vez, suficientemente justificado por esta misma experiencia. Con tales principios la razón se eleva cada vez más (como exige su propia naturaleza), llegando a condiciones progresivamente más remotas. Pero, advirtiendo que de esta forma su tarea ha de quedar inacabada, ya que las cuestiones nunca se agotan, se ve obligada a recurrir a principios que sobrepasan todo posible uso empírico y que parecen, no obstante, tan libres de sospecha que la misma razón ordinaria se halla de acuerdo con ellos. Es así como incurre en oscuridades y contradicciones. Y, aunque puede deducir que éstas se deben necesariamente a errores ocultos en algún lugar, no es capaz de detectarlos, ya que los principios que utiliza no reconocen contrastación empírica alguna por sobrepasar los límites de toda experiencia. El campo de batalla de todas estas inacabables disputas se llama *metafísica*."

consecución de un fin depende de ciertos factores que delimitan las condiciones de posibilidad de dicha acción (proyecto o práctica). Esto es tan simple como decir que para arrojar una piedra hace falta una piedra; o que, para encontrar el Santo Cáliz, éste ha de existir; o que, para escribir, es preciso saber leer. Las condiciones de posibilidad de una acción definen un marco, un espacio "dentro" del cual la acción es posible. "Fuera" de ese marco o espacio la acción no se puede ejecutar con éxito. Es fácil constatar que ciertas acciones son posibles en determinadas circunstancias (arrojar una piedra) y otras imposibles en toda circunstancia (dibujar un círculo cuadrado). El problema que presentan ciertos proyectos o prácticas filosóficas es que involucran acciones que resultan imposibles en cualquier circunstancia por razones muy peculiares. Una de ellas es que la ejecución de dichas acciones presenta problemas de *circularidad viciosa*, afecta a totalidades que deben estar definidas antes de emprender la acción (para que ésta tenga éxito), pero que sólo pueden estar plenamente definidas cuando la acción concluye, porque la acción *construye* uno de los elementos de dicha totalidad.

En una brevísima narración, Borges nos habla de unos cartógrafos que emprenden la tarea de construir un mapa singular del imperio que habitan. El mapa tiene las dimensiones del propio imperio y representa fielmente *todo* lo que en él podemos encontrar.[30] Cabe suponer que en un

[30] "... En aquel Imperio, el Arte de la Cartografía logró tal Perfección que el mapa de una sola Provincia ocupaba toda una Ciudad, y el mapa del Imperio, toda una Provincia. Con el tiempo, esos Mapas Desmesurados no satisficieron y los Colegios de Cartógrafos levantaron un Mapa del Imperio, que tenía el tamaño del Imperio y coincidía puntualmente con él. Menos Adictas al Estudio de la Cartografía, las Generaciones Siguientes entendieron que ese dilatado Mapa era Inútil y no sin Impiedad lo entregaron

momento determinado los cartógrafos se percataron de que su tarea era irrealizable. Para tener las propiedades deseadas, el mapa del imperio debería contener *una representación del propio mapa*, un mapa*. Pero, para que la representación del mapa fuese *exacta*, el mapa* debería ser cartografiado como una representación *exacta* del imperio que contuviese a su vez una nueva representación *exacta* del mapa, un mapa**. Es fácil adivinar que el proceso se repite *ad infinitum*. Para tener esperanzas de concluir un mapa como el mencionado, todo lo que hemos de representar debe estar a la vista en su totalidad, sin embargo, una de las cosas que hemos de representar es *la propia representación* del imperio. El mapa sólo se podría completar si ya existiese el mapa completo, el círculo vicioso es inevitable.

La imagen filosófica de las paradojas que hemos esbozado sostiene que ciertos proyectos filosóficos tienen forzosamente esta estructura circular, persiguen resultados tan generales que afectan al propio estatus de los proyectos que las persiguen. Ésta, y no otra, es la *generalidad* que nos preocupa y que antes mencionábamos. Algunos proyectos como, por ejemplo, el de demostrar si "todos los números naturales son divisibles por sí mismos y por la unidad" son generales porque persiguen establecer un hecho que afecta a la totalidad de elementos de un determinado universo de discurso: los números naturales. Esta generalidad, sin embargo, no es problemática porque

---

a las Inclemencias del Sol y de los Inviernos. En los desiertos del Oeste perduran despedazadas Ruinas del Mapa, habitadas por Animales y por Mendigos; en todo el País no hay otra reliquia de las Disciplinas Geográficas." J. L. Borges, "Del rigor de la ciencia", en *El hacedor*, Emecé, Buenos Aires, 1960.

nuestra demostración *no se ve afectada* por la naturaleza del hecho que persigue demostrar, la demostración es ajena a las consecuencias que se derivan de ella. Es habitual encontrar esta generalidad en las ciencias. La generalidad que nos (pre)ocupa aquí, en cambio, es la generalidad de toda acción que pretenda establecer la validez de un hecho para la totalidad de los elementos de un universo de discurso *que incluye a la propia acción.* Alguien que emprende una investigación para *conocer* si es posible el *conocimiento*, emprende una investigación para conocer, entre otras cosas, si la propia investigación que ha emprendido tiene posibilidades de éxito. Cuando una acción se dirige hacia una totalidad en la cual está incluida, se dirige, entre otras cosas, hacia sí misma (al igual que el mapa de Borges) y esto genera problemas de circularidad.

Pero la *circularidad* no es la única causante de las paradojas (aunque sea un ingrediente fundamental), las paradojas aparecen cuando *diagonalizamos* sobre proyectos definidos circularmente. Abordaremos esta cuestión en los capítulos 2 y 3, de momento esbozaré el problema con un ejemplo ilustrativo: Supongamos que, previendo las dificultades expuestas, los cartógrafos de Borges decidieron construir un mapa en el que se representó *todo* lo que hay en el imperio *excluyendo sólo* aquello que *se representase a sí mismo*. El nuevo mapa *podía representar mapas*, pero no mapas que se representasen a *sí mismos* evitando así bucles infinitos como los descritos. Los cartógrafos finalizaron su mapa pero, al hacerlo, se dieron cuenta, sin duda, de que habían fracasado. El mapa debía representar *todo objeto del imperio que no se representase a sí mismo*. El mapa *no* se representó a sí mismo de hecho, por tanto, *debió* haberse representado. Sin embargo, si se hubiese representado, entonces *no debería* haberlo hecho. De ese modo los cartógrafos comprendieron que no habían construido una representación de *todos los objetos del imperio que no se representan a sí mismos*. Y no lo hicieron porque tal proyecto era imposible. Quizá ésa fuese la razón íntima por la que decidieron abandonar su monstruoso proyecto.

Como hemos dicho, las preguntas que formula la filosofía no parecen detenerse en ninguna parte. Aunque toda investigación filosófica parta necesariamente de unos principios, no hay principios exentos, "*prima facie*", de una posible investigación filosófica. Pero, ¿qué ocurre cuando nuestra discusión sobre la validez o el estatus de unos principios ha de presuponer ineludiblemente los propios principios que discute, dada la generalidad de los mismos? ¿Cómo establecer o poner en duda la validez de algo mediante un argumento que necesita presuponer la validez de aquello que acaba estableciendo o poniendo en duda? La filosofía opera así a menudo, intenta establecer o justificar algo mediante métodos y argumentos que sólo pueden ser válidos si aquello que pretende justificar mediante ellos lo es ya de antemano. Ésta es la perplejidad en la que se encuentra al tratar de entender los conceptos más básicos: Cuando intentamos determinar cuáles son los "primeros principios" de *toda* investigación, lo hacemos mediante una investigación que ha de partir de *esos principios*. Cuando intentamos justificar la lógica (entendida en términos fregeanos como estudio de las leyes del pensamiento), lo hacemos razonando *a través de ella*. Cuando reflexionamos sobre la "posibilidad del conocimiento", lo hacemos en busca de *conocimiento*. Cuando nos preguntamos sobre "la posibilidad del significado", lo hacemos mediante un *lenguaje* cuya inteligibilidad no cuestionamos. Y así en infinidad de ejemplos.

Este método es extraño, lo que se pretende mediante una investigación o un razonamiento es descubrir o probar cómo es posible algo de lo que la propia investigación o razonamiento son un ejemplo. Y lo más curioso de todo es que a veces llegamos a la conclusión de que lo investigado (cuya verdad es, a su vez, presupuesta por la investigación para ser válida) *es imposible*. Demostrar que el conocimiento *es imposible* mediante una *investigación* encaminada a *obtener conocimiento* es como construir un mapa que represente sólo a todas aquellas cosas que no se representan a sí mismas, un proyecto imposible y contradictorio.

¿Cómo podríamos saber que es imposible saber? El método de Descartes nos pide, por ejemplo, que pongamos en duda todo aquello de lo que "podamos" dudar (incluso aquello de lo que, de hecho, no dudamos) y que sólo nos detengamos allá donde la duda roce lo "ininteligible": sólo así alcanzaremos un conocimiento cierto, claro y preciso. Pero, obviamente, nuestras dudas se *deben* volver ininteligibles llegados a cierto punto. Nuestras dudas se deben volver ininteligibles cuando cuestionan aquello en lo que se sustenta la validez del método, en caso contrario el propio método carecería de sentido. Para aceptar el *cogito*, Descartes debe aceptar primero el método, y con él sus criterios de claridad y distinción y las categorías que le permiten formular inteligiblemente tanto el *cogito* como el método. Pero, al mismo tiempo, el *cogito* se concibe como la piedra angular que justifica todo el conocimiento y, por tanto, el conocimiento del propio método, de sus criterios y de las categorías en que formula sus resultados. La filosofía se mueve a menudo en círculos de los que no puede escapar y, sin embargo, imagina la posibilidad de un espacio fuera de las órbitas que describe, un espacio cuya existencia nos parece inteligible pero cuyo acceso nos está vedado. Pensar en "ese espacio", se convierte en algo inevitable, pero forzosamente paradójico.

I EL ESQUEMA DE INCLUSIÓN Y LAS PARADOJAS REFLEXIVAS

2 La estructura de las paradojas reflexivas: el Esquema de Inclusión

2.1 Introducción

Hablar de paradojas sin más, prescindiendo de ejemplos o ulteriores precisiones del término 'paradoja', resulta siempre insatisfactorio. Pocas cosas con validez general acerca de ellas podríamos añadir a lo ya expuesto en las primeras secciones del capítulo anterior. Decíamos allí que una paradoja es una idea extraña u opuesta a la común opinión y al sentir de las personas, una idea que suele presentarse bajo la forma de un razonamiento (aparentemente) válido que, partiendo de premisas (aparentemente) verdaderas, alcanza una conclusión (aparentemente) falsa. Decíamos, también, que las paradojas o "antinomias" suponen un reto. Un razonamiento paradójico cuestiona la percepción u opinión que tenemos de la validez o eficacia de las capacidades cognitivas, principios y mecanismos epistémicos involucrados en la obtención de la conclusión que establece. En la medida en que aceptar dicho razonamiento en todos sus pasos nos resulte intolerable, debemos revisar en uno u otro sentido nuestras creencias con respecto al mismo. Como ya apuntamos, las reacciones frente al problema que suponen las paradojas son tan variadas como diferentes son las interpretaciones de su significación filosófica. Sin embargo, cobrar conciencia de este problema no nos ayuda a comprender mejor qué son las paradojas, qué forma tienen, cómo se originan o cuáles son sus causas fundamentales. Por este motivo, dejaremos momentáneamente de lado la relevancia que una u otra paradoja particular puedan tener para la filosofía y nos centraremos en el presente capítulo en una cuestión de carácter descriptivo: ¿Guardan las paradojas alguna relación o similitud relevante entre sí? ¿Cabe identificar una estructura subyacente común o un patrón recurrente que nos permita *describir* diferentes paradojas como ejemplos de un mismo problema?

Una respuesta afirmativa general a estas preguntas parece del todo improbable dado el elevado número de paradojas existentes y la diversidad de circunstancias, conceptos y razonamientos que las originan.[1] Sin embargo, quizá sea posible ofrecer una respuesta parcialmente positiva. Quizá sea lícito agrupar algunas paradojas de acuerdo con ciertas similitudes o rasgos relevantes y ofrecer una descripción estructural unificada de cada grupo. Esto ayudaría enormemente a simplificar la abrumadora diversidad de paradojas que encontramos en la literatura especializada, pero, sobre todo, nos permitiría identificar y entender mejor tanto las causas que originan una paradoja como los rasgos que hacen de paradojas pertenecientes a grupos diferentes fenómenos completamente distintos. La idea fundamental es, pues, la siguiente: dos paradojas diferentes (por apelar a conceptos, principios y hechos diferentes) plantearán habitualmente problemas distintos en sus respectivos ámbitos de origen. Sin embargo, por dispares que sean las circunstancias en las que surgen, si somos capaces de identificar una estructura interna común a ambas, entonces debe existir al

---

[1] La lista sería interminable, Sainsbury 1995 contiene quizá la colección más completa de ejemplos, pero faltan muchas de las paradojas que aquí discutiremos, incluidas algunas muy interesantes, como la de Yablo [13].

menos un nivel de abstracción desde el que ambas se puedan concebir como ejemplos de un mismo problema general.

En un libro reciente, *Beyond the Limits of Thought* (2002, en lo sucesivo: *BLT*),[2] Graham Priest asume la tarea de proporcionar una descripción estructural unificada de cierto número de paradojas. El presente capítulo contiene una discusión crítica (y, en gran medida, una defensa matizada) de las conclusiones alcanzadas por Priest con respecto a las preguntas que hemos formulado. Siguiendo la estrategia expositiva del propio Priest nos centraremos inicialmente en un grupo limitado de paradojas a las que sumaré nuevos ejemplos sobre los que poner a prueba sus afirmaciones.

2.2 La clasificación de las paradojas "reflexivas": Russell contra Ramsey

En un artículo de 1908, "Mathematical Logic as Based on the Theory of Types", Bertrand Russell reflexiona sobre las causas y rasgos comunes de ciertas paradojas que obstruyen el desarrollo de lo que él da en llamar "lógica matemática". Estas paradojas amenazan cualquier definición o concepción coherente de nociones tales como "verdad", "conjunto", "clase", "relación", "definible", "número", "cardinal" u "ordinal". Russell considera en el citado artículo la siguiente lista de paradojas (advirtiéndonos que no es en absoluto exhaustiva):[3] [1] La paradoja del mentiroso, [2] la paradoja de Russell, [3] una paradoja

[2] Y, con anterioridad, en Priest 1994a.
[3] Russell 1908 (pp. 59-60).

relacional, [4] la paradoja de Berry, [5] la paradoja de König, [6] la paradoja de Richard y [7] la paradoja de Burali-Forti.

Russell se refiere a estas paradojas mediante el nombre genérico de "paradojas de autorreferencia o reflexividad"[4] ya que para él todas ellas tienen algo en común. Todas estas contradicciones presuponen la existencia de una totalidad en términos de la cual podemos definir un nuevo elemento que, por definición, debe a la vez estar y no estar en dicha totalidad.[5] En la medida en que estas paradojas tienen un origen común, Russell propone también una solución única para todas ellas, su teoría de tipos basada en el "Principio de Circularidad Viciosa" (PCV) según el cual ninguna totalidad puede contener miembros definidos en términos de ella misma.[6]

Diecisiete años más tarde, Frank Ramsey considerará la misma lista de paradojas que reclamaron la atención de Russell añadiendo tan sólo una más, la paradoja de Grelling (que él atribuye a Weyl) [8].[7] Las conclusiones de Ramsey, sin embargo, difieren ampliamente de las de Russell. Ramsey evita utilizar un nombre genérico para todas estas paradojas y, lejos de considerarlas muestras de un mismo fenómeno, las divide en dos grupos, A y B. En el primero incluye la paradoja de Russell

---

[4] Russell, opus cit. p. 61.

[5] "In each contradiction something is said about *all* cases of some kind, and from what is said a new case seems to be generated, which both is and is not of the same kind as the cases of which *all* were concerned in what was said." Ibíd., p. 61. Más adelante, Russell dice: "Thus all our contradictions have in common the assumption of a totality such that, if it were legitimate, it would at once be enlarged by new members defined in terms of itself." Ibíd., p. 63.

[6] "These fallacies, as we saw, are to be avoided by what may be called the 'vicious-circle principle'; i.e., 'no totality can contain members defined in terms of itself'." Ibíd., p. 75.

[2], la de Burali-Forti [7] y la paradoja relacional [3]; en el segundo aparecen la paradoja del mentiroso [1] y las paradojas de König [5], Richard [6], Berry [4] y Grelling [8]. La división se justifica del siguiente modo:

> "The principle according to which I have divided them is of fundamental importance. Group A consists of contradictions which, were no provision made against them, would occur in a mathematical system itself. They involve only logical or mathematical terms such as class and number, and show that there must be something wrong with our logic or mathematics. But the contradictions of group B are not purely logical, and cannot be stated in logical terms alone; for they all contain some reference to thought, language, or symbolism, which are not formal but empirical terms. So they may be due not to faulty logic or mathematics, but to faulty ideas concerning thought and language. If so, they would not be relevant to mathematics or to logic, if by 'logic' we mean a symbolic system, though of course they would be relevant to logic in the sense of analysis of thought [*NOTA*: These two meanings of 'logic' are frequently confused. It should be clear that those who say mathematics is logic are not meaning by 'logic' at all the same thing as those who define logic as the analysis and criticism of thought.]"[8]

Comparando las clasificaciones de Russell y Ramsey, un primer detalle salta a la vista: la primera se basa en criterios de marcado carácter estructural –(1) la existencia de una totalidad X; y (2) la existencia de un elemento a de X definible sólo circular o autorreferentemente a partir de

---

[7] Ramsey 1925, p. 20.

[8] Opus cit., pp. 20-1. Antes que Ramsey, Peano (1906, p. 157) ya apunta esta división al señalar que sólo algunas de las paradojas que nos ocupan pertenecen a la matemática, las otras pertenecen a la lingüística.

consideraciones sobre *X*–, la clasificación de Ramsey, en cambio, se basa en la procedencia, el contenido y la naturaleza de los términos que aparecen en la formulación de las distintas paradojas. Según este criterio, el hecho de que una paradoja se formule exclusivamente en términos matemáticos o lógicos es relevante a la hora de clasificarla, calcular el alcance de sus consecuencias y buscar una solución.

Para hacer honor a la verdad, no está claro, a partir del texto citado, cuál es la mayor preocupación de Ramsey. Quizá Ramsey quiera advertirnos meramente de que las paradojas se deben abordar en primer lugar allá donde sus consecuencias son, por un lado, más devastadoras y, por otro, más fáciles de atajar, esto es, en los cálculos deductivos formales (o sistemas "lógicos"): teoría de conjuntos, lógica de primer orden, etcétera. Lo urgente es demostrar la coherencia de los términos empleados en estos sistemas formales. Abordar el problema de las paradojas en general sería una tarea sólo relevante para el proyecto más ambicioso, pero menos acuciante, de establecer analíticamente "las leyes del pensamiento" (la "Lógica" con mayúscula). También sería posible, por contra, que Ramsey estuviese fundamentalmente interesado en establecer una distinción entre las paradojas de tipo A y las paradojas de tipo B apoyándose en la naturaleza radicalmente distinta de los términos en que se formulan, los términos A (por abreviar) son "formales" y los B, "empíricos".

Tanto en un caso como en otro, lo que parece claro es que los criterios que baraja Ramsey en 1927 para trazar su distinción entre términos formales y empíricos son inválidos, padecen las limitaciones de su perspectiva histórica. A partir de Tarski, Gödel y otros sabemos que es posible formalizar y hablar con rigor de nociones tales como "verdad", "satisfacción" o "definibilidad" y que las teorías formales de modelos son parte fundamental de la matemática y la lógica modernas. Obviamente las nociones de verdad, etc. que usamos en estas teorías no coinciden con nuestra concepción ordinaria de "verdad", pero tampoco nuestra noción ordinaria de conjunto, clase o colección se corresponde con la que utilizamos en teoría de conjuntos. Ramsey se equivoca al pensar que ciertas nociones no pueden formar parte de un sistema deductivo formal. De hecho, es muy difícil trazar una distinción exitosa entre conceptos "puramente matemáticos o lógicos" y conceptos "empíricos" ya que los primeros son habitualmente abstracciones formales de conceptos ordinarios con un alto contenido empírico de los que, en cierto modo, son "parasitarios". No podemos afirmar con rotundidad qué términos de nuestro lenguaje son completamente ajenos a la formación de un cálculo formal riguroso, cualquier conjetura al respecto es susceptible de ser refutada con el paso del tiempo (pensemos en las afirmaciones que Kant hace en la *Crítica de la razón pura*, Bviii, sobre la lógica aristotélica). Desde esta perspectiva, las observaciones de Ramsey con respecto a la existencia de términos "lógicos" y términos "empíricos" no constituyen una base sólida para trazar una distinción profunda entre las paradojas de A y B.[9]

Sin embargo, que las razones aportadas por Ramsey para distinguir las paradojas A de las B sean erróneas, no significa que no sea

[9] Priest se apoya en este tipo de razones para rechazar la clasificación de Ramsey (*BLT*,

posible trazar dicha distinción sobre otra base. De hecho, hoy en día se asume implícitamente una dicotomía entre paradojas de la "teoría de conjuntos" (A) y paradojas "semánticas" (B) al abordar sus soluciones de forma independiente.[10] ¿Tiene esta distinción sentido?

En el presente trabajo deseamos reivindicar en lo esencial las posturas de Russell y Priest. Como dijimos al principio, la existencia de similitudes estructurales sugiere, al menos desde cierto nivel de abstracción, la presencia de un problema común. La regla de *modus ponens*, por ejemplo, se puede aplicar a enunciados empíricos o a enunciados de la aritmética, el contenido y la naturaleza de los términos empleados en estos enunciados es muy diferente, pero en ambos casos nos encontramos frente a la aplicación de una *única* regla de inferencia, ambos razonamientos tienen la *misma estructura*. Desde este punto de vista una descripción estructural unificada de las paradojas reflexivas revelaría que todas ellas son ejemplificaciones de un mismo problema de fondo. Ahora bien, ¿qué evidencia existe en este sentido? La descripción que aporta Russell de algunos rasgos estructurales presentes en las paradojas que nos ocupan parece acertada en muchos aspectos, pero resulta vaga. En el presente capítulo defenderemos que todas las paradojas expuestas hasta ahora, y otras que iremos viendo, encajan en lo que Priest llama "Esquema de Inclusión" [*Inclosure Scheme*], una

pp. 142-43).

[10] Un buen ejemplo del "triunfo" de la distinción de Ramsey es la naturalidad con la que Quine (1976b, pp. 10-11) la asume cuarenta años después (ya sin citar a Ramsey) al dividir las paradojas en dos "familias" (según los términos que involucren sean semánticos o matemáticos) y sugerir que deben solucionarse de forma diferente (ibíd. p. 11).

descripción estructural mucho más precisa. Defenderemos, también con Priest, que el Esquema de Inclusión ofrece criterios relevantes, no sólo para presentar de forma unificada las paradojas reflexivas, sino también para distinguir dentro de ellas, apelando una vez más a criterios estructurales, varias *sub*familias relevantes.

2.3 ¿Cuál es el problema?

Hemos hablado en todo momento de “paradojas”. Nos hemos preguntado, en particular, si ciertas paradojas comparten o no una misma estructura. Sin embargo, ahora deseamos corregir y precisar la formulación del problema. Lo que identificaremos aquí no es la estructura de ciertas paradojas sino, más bien, la estructura de cierto tipo de *razonamientos* cuya conclusión es una contradicción. Sostendremos que estos razonamientos aparecen en todas las paradojas que hemos considerado (y en otras muchas), pero también en las refutaciones legítimas de ciertas hipótesis o supuestos ilícitos. Llamaremos *contradicciones de inclusión* [*Inclosures*] a los razonamientos, sean o no paradójicos, que ejemplifiquen el patrón estructural (el *Esquema de Inclusión*) que describiremos más adelante. Esta modificación es, en el fondo, intuitiva. Pensemos en los siguientes razonamientos:

(A) “Si un ornitorrinco es un mamífero, entonces no pone huevos. Un ornitorrinco es un mamífero, luego no pone huevos.”

(B) "Si 2 es el sucesor de 1, entonces 2 $\neq$ 1. 2 es el sucesor de 1, luego 2 $\neq$ 1."

Ambos razonamientos tienen la misma forma, son aplicaciones correctas de la regla de *modus ponens* sobre sus respectivas premisas. Sin embargo, mientras que (B) establece la conclusión deseada, puesto que sus premisas son verdaderas, (A) no lo consigue al ser falsa la implicación de una de sus premisas. Una contradicción de inclusión inocua y una peligrosa se pueden distinguir de forma similar. En la primera, es posible detectar una incorrección en el establecimiento de la conclusión (una premisa falsa o una regla de inferencia inaceptable) siendo así posible convertir el razonamiento en una *reductio ad absurdum* de alguno de sus supuestos. En una paradoja de inclusión, en cambio, esto no es posible. Las premisas son verdaderas y las reglas de inferencia correctas, pero la conclusión sigue siendo una contradicción.

Como decíamos en el primer capítulo, una paradoja es algo que *percibimos* como ajeno a la común *opinión* y al *sentir* de las personas. Así pues, las paradojas dependen en gran medida de la opinión que nos merece la verdad de ciertos enunciados y la validez de ciertos métodos deductivos. En la medida en que dicha opinión está sometida a cambios y contingencias históricas, resulta siempre controvertido afirmar que algo es una paradoja. Nos enfrentamos aquí a un dilema. Por un lado, es inviable asumir la posible validez de una paradoja "genuina", i.e. de un razonamiento válido con premisas verdaderas y conclusión falsa. Por otro, la evidencia empírica de nuestro fracaso continuado a la hora de evitar satisfactoriamente la inferencia de ciertas contradicciones, nos empuja a pensar (o a temer) que, tal vez, existan paradojas "genuinas". Ahora bien, si éstas existen, parece claro que han de ser irresolubles, de lo contrario no habría paradojas. Toda paradoja resuelta se "deshace", con el tiempo, deja de percibirse con extrañeza y de existir, por tanto, como algo que va en *contra* –o que está *más allá*– (*para*) de la *opinión* (*doxa*). Nuestras limitaciones cognitivas suponen un serio obstáculo a la hora de decidir si hay o no paradojas "genuinas" (o cuáles pueden ser, de haberlas). Con todo, descubrir una forma común en los razonamientos que identificamos detrás de ciertas paradojas y también detrás de ciertas demostraciones ampliamente aceptadas (pero percibidas tal vez con extrañeza en su momento), nos ayudaría, no sólo a entender mejor cómo abordar el problema de las paradojas, sino también a comprender mejor cuál es su origen y a valorar qué relación guardan con otros fenómenos.

El problema que queremos resolver es, pues, el de *describir las contradicciones de inclusión*: cómo surgen, cuáles son sus elementos estructurales básicos y cómo se articulan. Hoy en día, algunas de las "paradojas" de inclusión consideradas por Russell y Ramsey, sobre todo aquellas expresables exclusivamente en términos de teoría de conjuntos, han perdido para muchos (aunque no para todos como veremos en capítulos sucesivos, especialmente 4 y 5) su carácter paradójico. Los razonamientos problemáticos se han reinterpretado como reducciones al absurdo de la existencia de ciertas totalidades, funciones, etc. supuestas por las premisas y se han reubicado en el contexto de pruebas válidas. No es nuestra intención discutir ahora hasta qué punto esta reconversión ha sido exitosa, pero sí examinar qué tienen en común todos estos argumentos, cómo se llega a la demostración de una contradicción en todos ellos. Que a partir de la contradicción podamos construir después una *reductio* o no es una cuestión diferente, lo que queremos saber es cómo llegamos a inferir una contradicción *de inclusión.*

Uno de los rasgos que se ha destacado como responsable de la contradicción en el caso de las paradojas (pero también en los teoremas) es la presencia de algún tipo de "*autorreferencia*" o "*reflexividad*" en el elemento central de la paradoja: la oración del mentiroso, el conjunto de Russell, la descripción 'el menor ordinal indescriptible en castellano', etcétera. Sin embargo, está claro que este rasgo no puede ser el único causante de la contradicción. Oraciones como (1) = 'Esta oración contiene cinco palabras'; descripciones como (2) = 'Las descripciones con menos de diez palabras'; predicados como (3) = 'es un predicado'; o conjuntos como el conjunto de todos los conjuntos con más de cinco elementos (que seguramente se pertenecerá a sí mismo por tener más de cinco elementos) no son expresiones particularmente problemáticas o contradictorias pese a ser claramente "autorreferentes" –(1), (2)– o "auto aplicativas" –(3)–.

Quizá el problema radique, como se ha sugerido también, en que la autorreferencia o reflexividad va ligada a cierta "*circularidad viciosa*" en los casos problemáticos. La evaluación semántica de (1) o (2), por ejemplo, no es circular, se apoya en un hecho *independiente* de la propia evaluación, i.e., el número de palabras en (1) y (2). Otro tanto ocurre con (3). Consideremos la oración (4) = '(3) es un predicado'. La verdad de (4) no depende de que (3) ocupe una posición predicativa *en* (4) sino de que ocupe una posición predicativa *en una oración cualquiera*. Que (3) sea o no un predicado *no* depende circularmente ni de la verdad de (4) ni de la función sintáctica que desempeña (3) en (4). Sin embargo, en los casos problemáticos *sí* hay circularidad. En el caso de la oración del mentiroso, μ = 'μ es falsa', la evaluación semántica de μ depende de sí misma. Para saber si μ es verdadera o falsa debemos saber *de antemano* si μ es verdadera o falsa. Algo parecido ocurre con 'heterológico', la verdad de (5) = "'heterológico' es heterológico" depende de la definición de 'heterológico', pero su aplicación concreta a este caso no nos ayuda porque *depende, a su vez*, de la evaluación semántica de (5). Del mismo modo, para especificar los miembros del conjunto de todos los conjuntos que no se pertenecen a sí mismos, R, necesitamos *antes* saber si R se pertenece o no a sí mismo, etcétera... Así pues, el problema parece radicar en la circularidad. Éste es, de hecho, el diagnóstico que hace Russell en 1908 al prohibir que "ninguna totalidad contenga miembros definidos en términos de ella misma", como ocurre con R, 'heterológico' o μ –si interpretamos μ, al igual que Russell, como la oración que dice de sí misma que *no pertenece a la totalidad de las oraciones verdaderas*–.

El "Principio de Circularidad Viciosa" (PCV) russelliano, sin embargo, no es del todo satisfactorio, pues, tal y como está formulado, excluye la posibilidad de objetos y definiciones totalmente inofensivas. Pensemos en "el mejor jugador de la selección argentina de fútbol en 1986" (al parecer, el contraejemplo original es de Ramsey y no muy diferente). Hablamos de Maradona, por supuesto, pero (en contra del precepto russelliano) lo hemos descrito, sin ningún problema, a partir de la totalidad a la que perteneció como jugador en 1986: la selección argentina de fútbol. La importancia del contraejemplo de Ramsey es, no obstante, secundaria. Con toda seguridad podemos determinar los miembros del equipo argentino de 1986 con independencia de quién fuese su mejor jugador, de igual modo, podemos seleccionar o describir a Maradona sin tener en cuenta sus habilidades o su pertenencia a ningún equipo. Sólo tenemos un problema de circularidad cuando pensamos en totalidades que contienen *necesariamente* miembros que *sólo* se pueden definir o especificar a partir de la totalidad mencionada. Éstas son las totalidades que aparecen en las paradojas descritas y que debe tener en cuenta PCV. Si *sólo* pudiésemos identificar a *cada* miembro de la selección argentina de fútbol diciendo de él que es: "aquel que juega al fútbol en el equipo nacional argentino junto a otros 10 jugadores distintos", entonces tendríamos un problema de circularidad. Nuestro interés se centra exclusivamente en casos en los que aparece una entidad, A, que *sólo* puede definirse a partir de la definición previa de otra entidad, B, cuya definición depende, a su vez, de la propia definición de A. *Sólo* estos casos cuentan como círculos viciosos.

Pero, aun cuando podamos salvar el principio russelliano, cabe preguntarse si podemos rechazar en general la circularidad. Hay enunciados de carácter normativo cuya evaluación semántica parece depender circularmente de las normas que ellos establecen sin que ello sea percibido como un problema. Pensemos en alguien que establece por primera vez la norma de ajedrez: "Los alfiles se mueven en diagonal por el tablero". La verdad de este enunciado parece depender fundamentalmente del propio enunciado (cuando es proferido exitosamente con fines normativos).[11] Por otra parte, la idea de conjuntos no fundamentados (o hiper-conjuntos) cuya definición entraña círculos viciosos o regresos al infinito está ampliamanete aceptada hoy en día. Todos los conjuntos de teorías clásicas como ZF se pueden construir a partir de ∅ (y ω) mediante aplicaciones de operaciones conjuntistas. La clase resultante es la "jerarquía acumulativa". Esto no ocurre con los conjuntos no fundamentados. Un conjunto no fundamentado, *a*, podría pertenecerse a sí mismo o a algún elemento, *b*, necesario para la construcción de *a*; alternativamente, *a* podría depender de la construcción previa de una serie infinita de conjuntos no generados a partir de ∅. En estos casos la definición de *a* no se puede construir sólo a partir de ∅, ya que descansa sobre un círculo vicioso o un regreso al infinito.

Pero olvidémonos ahora de estos casos. Incluso si la circularidad fuese un problema y pudiésemos (y debiésemos) prescindir de los conjuntos no fundamentados y dar cuenta de la validez de enunciados normativos excluyendo explicaciones circulares, eso no nos ayudaría a delimitar mejor las contradicciones de inclusión. En estos razonamientos, la circularidad *no es el único rasgo estructural determinante* en la

[11] En realidad este ejemplo es dudoso. Los enunciados normativos pueden identificarse con enunciados imperativos u órdenes, que *carecen* de condiciones de verdad. Un ejemplo más claro de circularidad semántica forzosa se puede encontrar en ciertas definiciones de términos que no designan entidades de carácter exclusivamente normativo. Es de suponer que toda definición es verdadera o falsa (i.e., se cumple o no se cumple de lo definido). Esto significa que cualquier definición del concepto "verdad" habrá de ser a su vez verdadera o falsa, con lo que presuponemos circularmente aquello

inferencia de una contradicción. Pensemos en la oración ν = 'ν es verdadera'. La evaluación semántica de ν es tan circular como lo pueda ser la de μ, sin embargo, los razonamientos y problemas que asociamos a μ coinciden sólo parcialmente con aquellos que asociamos a ν. μ no puede ser evaluada coherentemente, ν, en cambio, sí. Del mismo modo, se puede demostrar que la noción de conjunto no fundamentado es tan coherente como lo pueda ser la noción de conjunto fundamentado. No todas las colecciones no fundamentadas se comportan como R, la clase de Russell. Forti, Honseell, Aczel, Barwise y Moss han elaborado demostraciones de consistencia relativa para diferentes teorías de conjuntos no fundamentados.[12] Así pues, los problemas que plantean las contradicciones de inclusión y los problemas de circularidad son tan similares y tan distintos como lo puedan ser μ y ν o como lo puedan ser R y cualquier conjunto no fundamentado coherente.

La circularidad es una *condición necesaria, pero no suficiente*, para que se den contradicciones de inclusión. El papel que juegan la negación y las totalidades mencionadas por Russell en su diagnóstico de las paradojas (la colección de todas las verdades, de todos los conjuntos, de todos los ordinales, etc.) parecen tan relevantes en la descripción de

---

que queremos definir. Discuto esta idea en J. Valor 2002 a partir de textos de Frege (1918, pp. 52-3) y Soames (1999, 21-9, 46-9).

[12] Para más detalles véase Barwise y Moss 1996. En el capítulo 9 ofrecen una demostración de consistencia relativa para ZFA. Demuestran que si ZFC es consistente, también lo es ZFA, y viceversa. 'ZFA' es el nombre que dan a ZFC (la presentación axiomática estándar de la teoría de conjuntos de Zermelo-Fraenkel añadiendo urelementos y el Axioma de Plenitud Fuerte), tras reemplazar el Axioma de Fundamentación por el Axioma de Anti-Fundamentación (AFA). Véase también Aczel 1988 o Devlin 1993, donde se prescinde de urelementos.

las contradicciones de inclusión como la circularidad viciosa que, sin duda, encierran algunos de los elementos del razonamiento. Pero, ¿cómo encajan en el puzzle todas estas piezas?

2.4 El esquema de Russell

Russell descubrió que muchas de las paradojas que podíamos describir en teoría de conjuntos tenían una estructura común. En 1905 ofreció una descripción general en la que encajaban tres de ellas: la de Russell [2], la de Burali-Forti [7] y la de Mirimanoff [9] sobre la jerarquía bien fundamentada de conjuntos. Éstas son las observaciones (bastante crípticas) de Russell con respecto a lo que todas estas paradojas comparten:

> Given a property φ and a function δ, such that, if φ belongs to all members of X, $\delta(X)$ always exists, has the property φ, and is not a member of X; then the supposition that there is a class Ω of all terms having property φ and that $\delta(\Omega)$ exists leads to the conclusion that $\delta(\Omega)$ both has and has not the property φ.[13]

En la "interpretación" que hace Priest de las palabras de Russell, la descripción anterior se representa mediante un esquema.[14]

[13] Russell 1905b, p. 142. La cita está extraída de *BLT* (p. 129), donde la notación original se ha adaptado a la utilizada por Priest. También en Priest 1994a.
[14] *BLT* (p. 129). También en Priest 1994a.

Esquema de Russell (ER):

(1) $\Omega = \{y : \varphi(y)\}$ existe.	*Existencia*
(2) Dada una función δ: $\wp(\Omega) \to \Omega$, si $X \subseteq \Omega$, entonces	
(a) $\delta(X) \notin X$	*Trascendencia*
(b) $\delta(X) \in \Omega$	*Clausura*

Según este esquema, las paradojas [2], [7] y [9] se originan cuando asumimos la existencia de una clase Ω (cuyos elementos satisfacen una propiedad φ) y de una función δ de los subconjuntos de Ω a Ω (i.e., δ: $\wp(\Omega) \to \Omega$) tal que la imagen de cualquier $X \subseteq \Omega$ satisface dos condiciones: $\delta(X) \notin X$ y $\delta(X) \in \Omega$. Priest llama respectivamente *trascendencia* y *clausura* a estas dos condiciones, ya que $\delta(X)$ está siempre *dentro de* Ω (*clausura*), pero *fuera de* X (*trascendencia*). La paradoja se obtiene cuando δ se aplica a Ω (uno de los subconjuntos de Ω), tenemos entonces una contradicción $\delta(\Omega) \notin \Omega$ y $\delta(\Omega) \in \Omega$.

Pero veamos como funciona el esquema mediante algunos ejemplos. Comencemos por [2]. Russell consideró que [2] era una versión depurada de una paradoja atribuible a Cantor [10].[15] Bajo este esquema, [2] se puede reescribir del siguiente modo: sea φ igual a la propiedad "$y \notin y$" (Ω es, pues, el conjunto de todos los conjuntos que no

[15] Se menciona este hecho en Priest 1994a y *BLT* (pp. 128-29), donde se muestra en qué sentido esto es así.

son miembros de sí mismos) y sea δ igual a la función de identidad sobre Ω,[16] tenemos entonces que:

1) $\Omega = \{y : y \notin y\}$ existe.

2) Si $X \subseteq \Omega$, entonces: $\delta(X) = id(X) = X$.

Tomemos un $X \subseteq \Omega$ y demostremos las condiciones de *trascendencia* (2a) y *clausura* (2b):

(2a) ***Trascendencia.*** Supongamos (para una *reductio*) que $\delta(X) \in X$:

$\delta(X) \in X$	$\Rightarrow$	$X \in X$	(definición de δ como '= ')
	$\Rightarrow$	$X \in \Omega$	($X \in X$ y $X \subseteq \Omega$)
	$\Rightarrow$	$X \notin X$	(definición de Ω)
	$\Rightarrow$	$\delta(X) \notin X$	(definición de δ como '= ')

(2b) ***Clausura.*** (Hemos demostrado "$\delta(X) \notin X$" por *reductio* sobre $\delta(X) \in X$):

[16] Según el esquema, δ ha de ser una función de $\wp(\Omega)$ a Ω, sin embargo, δ es aquí ¡la identidad sobre Ω! Esto es posible porque, si $\Omega = \{y : y \notin y\}$, podemos demostrar en ZF que $\wp(\Omega) = \Omega$. Brevemente:
(1) Supongamos que $x \in \wp(\Omega)$ (i.e., $x \subseteq \Omega$). Existen dos posibilidades: o $x \in x$, o $x \notin x$. Si $x \in x$, entonces $x \in \Omega$ (dado que $x \subseteq \Omega$); si $x \notin x$, entonces $x \in \Omega$ (por definición de Ω). Así pues, concluimos: $\wp(\Omega) \subseteq \Omega$.
(2) Ahora queremos probar que $\Omega \subseteq \wp(\Omega)$. Supongamos que $x \in \Omega$ (i.e., $x \notin x$), entonces o $x = \varnothing$ o $x \neq \varnothing$. Si $x = \varnothing$, entonces $x \in \wp(\Omega)$. Si $x \neq \varnothing$, entonces existe(n) un(os) conjunto(s) y_0, $(y_1, y_2, ...)$ tal(es) que $x = \{y_0, (y_1, y_2, ...)\} \in \Omega$. Por el axioma de fundamentación sabemos que $y_0, y_1, y_2, ... \in \Omega$ (i.e., no se pertenecen a sí mismos), así pues, $x \subseteq \Omega$ o, lo que es lo mismo, $x \in \wp(\Omega)$. Por lo tanto, $\wp(\Omega) = \Omega$. (Esto pone de

$$
\begin{array}{llll}
\delta(X) \notin X & \Rightarrow & X \notin X & \text{(definición de } \delta \text{ como `$=$')} \\
& \Rightarrow & X \in \Omega & \text{(definición de } \Omega) \\
& \Rightarrow & \delta(X) \in \Omega & \text{(definición de } \delta \text{ como `$=$')}
\end{array}
$$

La contradicción aparece cuando consideramos $\delta(\Omega)$, ya que entonces inferimos $\delta(\Omega) \notin \Omega$ y $\delta(\Omega) \in \Omega$.

De forma similar, podemos expresar mediante este esquema las paradojas de Burali-Forti [7], Mirimanoff [9] y Cantor [10].[17] Esto aporta sin duda evidencia a favor de la unidad estructural de estas paradojas y de la existencia de un único problema general de fondo. Para ser exactos, lo que encaja en el esquema no es la "paradoja" sino el *razonamiento* que conduce a una contradicción a partir de ciertas premisas. Que entendamos dicho razonamiento o no como una paradoja depende de nuestras intuiciones con respecto a la verdad de sus premisas. La paradoja sólo surge cuando las premisas parecen incuestionables. Si éste no es el caso, el razonamiento se convierte en un fragmento dentro de un razonamiento más amplio por *reductio ad absurdum* destinado a refutar alguna de las premisas. En estos casos se niega la existencia de Ω, o de δ, o de la propiedad φ, o de cualquier cosa sobre la que descanse la inferencia de una contradicción. Así se resuelven tradicionalmente las cuatro paradojas conjuntistas mencionadas (véase 4.4).

---

relieve el paralelismo que Russell encuentra entre su propia paradoja [2] y la de Cantor [10], donde, dado el conjunto de todos los conjuntos puros: V, tenemos $\wp(V) = V$).

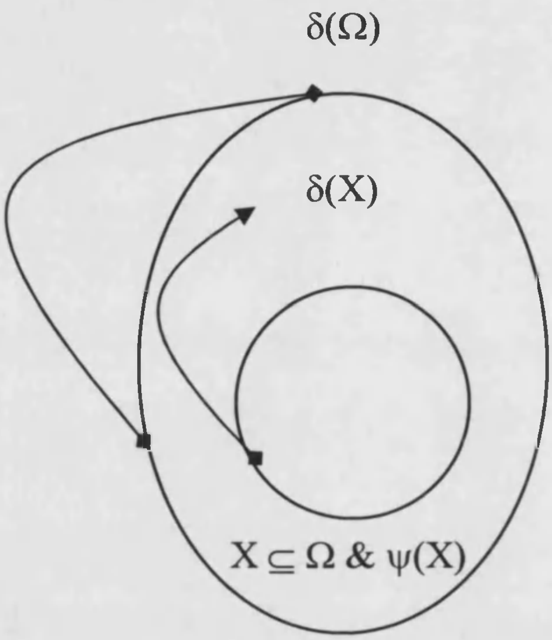
[17] Podemos encontrar una formulación explícita de estas paradojas bajo el mencionado esquema en Priest 1994a y *BLT* (capítulo 9). Véase también el próximo capítulo, 3.2.7.

2.5 El Esquema de Inclusión [*Inclosure Scheme*]

Como vimos, Russell estaba convencido de que todas las paradojas reflexivas tenían un origen común. Trató de identificar y bloquear las fuentes del problema al establecer el "principio de circularidad viciosa" (PCV), pero no consiguió su propósito. No todas las entidades que transgreden PCV son incoherentes o problemáticas en el sentido en que lo son los elementos centrales de las "paradojas reflexivas" (contradicciones de inclusión). Por consiguiente, PCV no aísla satisfactoriamente el problema relevante. La descripción estructural unificada que Russell proporcionó para las paradojas de Cantor [10], Russell [2], Burali-Forti [7] y Mirimanoff [9], se acerca mucho más a lo que buscamos pero carece de la generalidad apropiada. Cabe, pues, preguntarse si es posible encontrar una descripción estructural unificada de *todas* las paradojas reflexivas, incluidas aquellas que contienen términos semánticos tales como "verdad", "definible" o "descriptible".

Siguiendo los pasos de Russell, Priest (*BLT*, capítulos 9 y 10) ofrece una respuesta afirmativa a esta pregunta. De hecho, según Priest, Russell se acercó muchísimo a la solución definitiva. Modificando ligeramente el Esquema de Russell es posible obtener una descripción estructural de *todas* las paradojas consideradas por Russell y Ramsey (y de muchas otras). El esquema de Russell puede concebirse, de hecho, como un caso particular de un esquema más general que Priest bautiza con el nombre de [*Inclosure Scheme*] *Esquema de Inclusión* (EI). (Señalo en **negrita** las variaciones que introduce EI en ER.)

Esquema de Inclusión (EI):



$\Omega = \{y \colon \varphi(y)\} \,\&\, \psi(\Omega)$

(1) $\Omega = \{y : \varphi(y)\}$ existe **y** $\boldsymbol{\psi(\Omega)}$ (1) *Existencia*

(2) [Dada una función δ: $\mathbf{Y} \rightarrow \Omega$, donde $\mathbf{Y} \subseteq \wp(\boldsymbol{\Omega})$]
Si $X \subseteq \Omega$ **y** $\boldsymbol{\psi(X)}$, entonces

(a) $\delta(X) \notin X$ (2a) *Trascendencia*
(b) $\delta(X) \in \Omega$ (2b) *Clausura*

Contradicción: $\delta(\Omega) \notin \Omega$ y $\delta(\Omega) \in \Omega$

Al igual que en el Esquema de Russell (ER), δ se puede concebir como una función que "trasciende" siempre la totalidad X sobre la que se define, pero que se mantiene siempre "dentro de" los límites de Ω, de ahí los términos trascendencia y clausura. Una vez más, la contradicción se obtiene cuando δ toma por argumento Ω, tenemos entonces $\delta(\Omega) \notin \Omega$ y

$\delta(\Omega) \in \Omega$. [En el gráfico, cuando δ toma como argumento un $X \neq \Omega$, su valor es *un elemento de* Ω (*clausura*) que *no está en* X (*trascendencia*); cuando el argumento de δ es Ω, entonces el valor de δ se ubica, en palabras de Priest, en el "límite de Ω", esto es, "fuera" y "dentro" de Ω].

Priest muestra que todas las paradojas "semánticas" y "conjuntistas" consideradas por Ramsey encajan en este esquema. Es fácil ver, además, en qué sentido EI es una generalización de ER. En ER, δ es una función de $\wp(\Omega)$ a Ω. *Todos* los subconjuntos de Ω son relevantes para definir δ. En EI, por contra, δ es una función de $Y \subseteq \wp(\Omega)$ a Ω. *Sólo* necesitamos considerar una familia de subconjuntos de Ω: aquellos que satisfacen cierta propiedad ψ. ψ puede ser tan amplia (o estar tan restringida) como lo permita el esquema. Si tomamos como ψ una propiedad satisfecha por todos los subconjuntos de Ω (por ejemplo, la de "ser idéntico a sí mismo"), entonces obtenemos ER a partir de EI, ya que $Y = \wp(\Omega)$. En el otro extremo se encuentran las paradojas que sólo encajan en EI cuando tomamos ψ como la propiedad "ser igual a Ω", esto es, $Y = \{\Omega\}$. En este caso límite, sólo cuenta *un* subconjunto de Ω en la definición de δ, el propio Ω. Queda claro que cualquier definición de ψ ha de ser satisfecha por Ω: $\Omega \in Y \subseteq \wp(\Omega)$. Todas las paradojas consideradas por Russell en ER son, pues, casos particulares de EI en los que tomamos ψ como la propiedad universal "ser idéntico a sí mismo".

Pero, ¿por qué necesitamos modificar el Esquema de Russell? Priest muestra a través de un ejemplo en qué sentido ER no basta para dar una descripción estructural unificada de las paradojas. Consideremos la paradoja de König [5] sobre “el menor ordinal indefinible”.[18] Priest ofrece la siguiente interpretación de la misma en términos de EI: Siendo φ igual a “ser un ordinal definible” y ψ igual a “es definible”, tenemos que:

(1) $\Omega = \{y : y$ es un ordinal definible $\}$ existe **y Ω es definible.**

(2) Si $X \subseteq \Omega$ **y X es definible**, entonces: $\delta(X)$ = m.o.n.(X) (‘m.o.n.(X)’ abrevia ‘el menor ordinal que no está en X’).

En primer lugar, $\delta(X)$ siempre existe porque todo conjunto de ordinales tiene un menor elemento. On – X es un conjunto de ordinales (On es la colección de *todos* los ordinales) y tiene, por tanto, un menor elemento, a saber, m.o.n(X). Por otra parte, se puede comprobar fácilmente que δ satisface las condiciones de *trascendencia* y *clausura*. Obviamente $\delta(X) \notin X$, ya que $\delta(X)$ es el menor ordinal que *no está en* X. Ahora bien, dado que X es definible, también m.o.n.(X) lo es y, por tanto, $\delta(X) \in \Omega$. La paradoja aparece, como siempre, cuando consideramos $\delta(\Omega)$, ya que $\delta(\Omega) \notin \Omega$ y $\delta(\Omega) \in \Omega$.

Obsérvese que la descripción definida ‘m.o.n.(X)’ proporciona una definición (descripción) del menor ordinal fuera de X si, y sólo si, *X es definible* (*descriptible*). Si X (y, en particular, Ω) no fuese definible

[18] Quizá no esté claro que esta descripción proporcione una definición del citado ordinal. La paradoja se puede reformular sustituyendo ‘definible’ por ‘descriptible’ en el argumento. De este modo, como apunta Priest, la paradoja resulta menos controvertida y conserva su mordiente. En cualquier caso, por “definible”, se entiende aquí

(descriptible), entonces no podríamos completar la expresión 'm.o.n.(...)' con expresiones que denotasen, definiesen o describiesen a X u Ω y entonces ni m.o.n.(X) ni m.o.n.(Ω) serían definibles. El Esquema de Russell *no garantiza* que *todos* los subconjuntos de Ω sean definibles, ésa es la razón por la que necesitamos añadir una restricción adicional, la que introduce ψ en el esquema modificado. Se puede comprobar que este problema se reproduce en todas las contradicciones de inclusión que no encajan en el Esquema de Russell, por eso necesitamos el Esquema de Inclusión.

En los casos en los que ψ no es trivial (ψ es trivial cuando ψ = "*ser idéntico a sí mismo*"), la misión de ψ es especificar una propiedad que permita seleccionar una colección de subconjuntos de Ω mediante una característica compartida que sea relevante para la formación de la contradicción de inclusión. Una de las más recurrentes es la propiedad de ser "definible o expresable lingüísticamente" bajo ciertos parámetros ("en el lenguaje *L*", "con menos de *x* sílabas", etc.). Sin embargo, en algunos casos importantes, la naturaleza de la colección Ω que interviene en la paradoja hace que el único subconjunto relevante en la formación de la contradicción de clausura sea Ω mismo. Priest no ofrece ejemplos claros de estos casos (véase 3.2.10), pero examinaremos alguno en breve.

Una de las ventajas del Esquema de Inclusión es que ofrece criterios estructurales para mostrar de un modo unificado diferentes contradicciones de inclusión, pero también para distinguir dentro de ellas

---

"descriptible": "Call something *definable* if there is some non-indexical noun-phrase (of English) that refers to it." (*BLT*, p. 131.)

diferentes subfamilias (nos ocuparemos de estas cuestiones con detenimiento en el próximo capítulo). Lo que Ramsey llamó paradojas de la teoría de conjuntos, siguiendo criterios no muy precisos, puede caracterizarse ahora como contradicciones de inclusión en las que ψ es una propiedad universal. Priest llama a estas paradojas "contradicciones en los límites de la de iteración" porque en ellas lo fundamental es que δ se puede concebir como un "generador de totalidades infinitas, Ω" (véase 3.2.7). En el resto de contradicciones, ψ juega un papel sustancial y las cosas son diferentes (véase 3.2.9-10). Priest parece sugerir que, en estos casos, no sólo es fundamental la totalidad Ω y el modo en que δ contribuye a generarla, sino también el hecho de que dicha totalidad guarde relación con limitaciones de nuestro lenguaje ("lo descriptible", "lo definible") o de nuestras capacidades cognitivas, imaginativas, etc ("lo verdadero", "lo cognoscible"). Este es el motivo por el que Priest dice de estas otras contradicciones de inclusión que se producen en "el límite de la expresión" o en "el límite de la cognición". Estas contradicciones no se diferencian entre sí según criterios estructurales, sino más bien semánticos. Parece, así, después de todo, que Priest vuelve a introducir por la puerta de atrás criterios no estructurales relacionados con el tipo de propiedad que es ψ, tal y como hizo Ramsey. Esto es en parte cierto, pero sólo en parte. Es innegable que Priest, a diferencia de Ramsey, logra dos objetivos: 1) identificar un esquema general, un esquema que, más allá de cualquier subdivisión ulterior, *unifica todas* las paradojas relevantes; y 2) ofrecer criterios *estructurales* para distinguir subfamilias de paradojas reflexivas sobre la base del conjunto $Y \subseteq \wp(\Omega)$

que determina ψ en EI. Existen más o menos tres posibilidades: o $Y = \wp(\Omega)$ (i.e., ψ es una propiedad universal); o $Y = \{\Omega\}$ (i.e., ψ es "ser idéntico a Ω"); o $Y \subset \wp(\Omega)$ y $\Omega \in Y$ (i.e., ψ engloba a una familia de subconjuntos de Ω que incluye a Ω). Hemos visto ejemplos de contradicciones de inclusión en "los límites de la iteración" [$Y = \wp(\Omega)$]: Russell [2], Burali-Forti [7] y Mirimanoff [9]. Hemos visto ejemplos en "los límites de la expresión" [$Y \subset \wp(\Omega)$ y $\Omega \in Y$]: König [5] (también Berry [4], Richard [6] y lo que Priest denomina: "paradoja de Berkeley" –véase *BLT*, pp. 69-70; 132-33). Veremos a continuación contradicciones en "los límites de la cognición" [$Y \subset \wp(\Omega)$ y $\Omega \in Y$]: el mentiroso [1] y Grelling [8]. Finalmente, veremos contradicciones de inclusión del tipo [$Y = \{\Omega\}$]. Priest no ofrece ejemplos simples, sólo ejemplos relacionados con peculiaridades conceptuales que muestran ciertas teorías filosóficas. Nosotros mostraremos aquí dos ejemplos más sencillos: la paradoja de la cita de Tarski [11] y la contradicción demostrada en el teorema de Cantor.

2.6 La "paradoja" de la cita de Tarski

En su famoso artículo "The Concept of Truth in Formalized Languages", Tarski (1933, pp. 159-62) defiende que las comillas no son expresiones complejas desde un punto de vista sintáctico, sino que deben entenderse como simples caracteres en los nombres en los que aparecen. Esta interpretación de las comillas impide interpretarlas como funciones que asignan a una expresión cualquiera de nuestro lenguaje un nombre de dicha expresión en el propio lenguaje. Entre las razones que apoyan la afirmación de Tarski encontramos una muy poderosa: Si consideramos las comillas como funciones que asocian nombres a expresiones, podemos derivar una paradoja semántica prescindiendo del concepto de verdad. Llamaremos a esta paradoja, paradoja de la cita de Tarski [11].

Pensemos en un lenguaje en el que podamos construir la oración autorreferente:

(c) $\forall p\ (c = `p' \rightarrow \neg p)$

i.e., 'c' es un nombre de la oración "$\forall p\ (c = `p' \rightarrow \neg p)$". Si se cumple, para cualesquiera dos oraciones P y Q, que 'P' = 'Q' $\rightarrow$ (P $\leftrightarrow$ Q) es un principio válido (como parece serlo si las comillas son, de hecho, una función), demostramos una contradicción demostrada. He aquí el razonamiento:[19]

1. $c = `\forall p\ (c = `p' \rightarrow \neg p)'$	(Premisa 1)
2. $`P' = `Q' \rightarrow (P \leftrightarrow Q)$	(Premisa 2)
3. $\forall p\ (c = `p' \rightarrow \neg p)$	(Suposición)
4. $c = `\forall p\ (c = `p' \rightarrow \neg p)' \rightarrow \neg \forall p\ (c = `p' \rightarrow \neg p)$	(instancia de 3)
5. $\neg \forall p\ (c = `p' \rightarrow \neg p)$	(Modus Ponens: 1, 4)
6. $\neg \forall p\ (c = `p' \rightarrow \neg p)$	(absurdo: 3-5)
7. $\exists p \neg (c = `p' \rightarrow \neg p)$	($\neg\forall \Leftrightarrow \exists\neg$: 6)
8. $\exists p\ (c = `p' \wedge p)$	(Interdef. $\rightarrow$, $\wedge$: 7)
9. $c = `q' \wedge q$	(Sup. -prueba por casos $\exists$ en 8)
10. $`q' = c = `\forall p\ (c = `p' \rightarrow \neg p)'$	(Def. '=': 1, 9)
11. $q \leftrightarrow \forall p\ (c = `p' \rightarrow \neg p)$	(Modus Ponens: 2, 10)
12. $\forall p\ (c = `p' \rightarrow \neg p)$	(Mod. Pon: 9 (q), 11 ($q \rightarrow$ 12))
13. $\forall p\ (c = `p' \rightarrow \neg p)$	(Eliminación $\exists$: 9-12)
14. Contradicción 13, 6	

[19] Las suposiciones y sus cancelaciones se representan mediante tabulaciones.

Aunque las comillas no fuesen funciones, como defiende Tarski, esta paradoja continuaría siendo inquietante. Si no fuésemos capaces de resolverla, mostraría que no hay una manera coherente de expresar (en ciertos lenguajes) un signo funcional que, al adherirse (como prefijo, sufijo, etc.) a una expresión, produjese un nombre de la misma.[20] No pretendemos buscar aquí soluciones para esta paradoja, tampoco examinar sus consecuencias, nuestro interés se centra en ver si la "paradoja" de la cita encaja en EI.

Un primer intento nos llevaría intuitivamente a leer c como una oración que dice de sí misma que "no es verdadera" o, para ser más exactos, que "si c es el nombre de alguna oración, dicha oración es falsa". Consideremos las siguientes oraciones:

(a) $a = b \rightarrow \neg$ verdadera(a). (d) $d = d \rightarrow \neg$ verdadera(d).

(b) $a = b \rightarrow \neg$ verdadera(a). (e) $\forall x\, (e = x \rightarrow \neg$ verdadera$(x))$.

Las oraciones a/b, d y e son paradójicas y mantienen cierta semejanza con c –especialmente e. La diferencia más relevante, aparte de la ausencia de cuantificadores en la mayoría de ellas, es que todas contienen el predicado 'verdadero'. Si consideramos como la versión más simple de la paradoja del mentiroso [1] la oración μ = '$\neg$ verdadera(μ)', está claro que todas estas oraciones se pueden presentar como versiones más o menos sofisticadas de esta paradoja. Con independencia de la versión que escojamos, se puede mostrar que esta paradoja encaja en EI (siempre que la oración μ relevante diga de sí misma que no es verdadera). En este caso tenemos φ = "es verdadera" y ψ = "es definible":

1) Ω = T = {y : verdadera(y) (i.e., y es verdadera)} existe **y T es definible.**

2) Si X $\subseteq$ T **y X es definible**: δ(X) = μ.

Dependiendo de la oración del mentiroso que queramos construir consideramos: μ = '$\mu \notin$ X'; o μ = '$\forall y$ ($\mu = y \rightarrow y \notin$ X)'; etcétera. En este contexto, "$y \in$ T" se puede interpretar como "y es verdadera" ("$y \notin$ T" será, pues, "y no es verdadera"). 'μ' es, por tanto, un nombre de la oración que dice de sí misma que no pertenece a X (o que, si algo se llama 'μ', entonces no pertenece a X; etcétera). Obviamente, cuando sustituimos X por T, se obtiene la oración del mentiroso: μ = '$\mu \notin$ T'; o μ = '$\forall$y (μ = y $\rightarrow$ y $\notin$ T)'; etcétera. Veamos si δ satisface las condiciones de trascendencia y clausura:[21]

[20] Véase, por ejemplo, S. Haack 1978 (pp. 172-5) y S. Soames 1999 (pp. 86-92). En ambos lugares se pueden encontrar exposiciones más detalladas del problema y referencias bibliográficas.

[21] Ofrecemos una prueba adaptada al mentiroso simple: μ_1 = '$\mu_1 \notin$ T'. Si añadimos las líneas ***en negrita***, obtenemos una prueba para la versión más compleja: μ_2 = '$\forall y$ ($\mu_2 = y \rightarrow y \notin$ T)'. En el caso de μ_2, podemos prescindir del principio: 'P' = 'Q' $\rightarrow$ (P $\leftrightarrow$ Q). Sólo necesitamos el principio de *Intercambiabilidad Salva Veritate de los Idénticos* (ISVI): a = b $\rightarrow$ (P(a) $\leftrightarrow$ P(b)), para demostrar las condiciones de trascendencia y clausura. Tanto en μ_1 como en μ_2, necesitamos el Esquema-T tarskiano: "$p \leftrightarrow x \in$ T" (i.e., $p \leftrightarrow x$ es verdadera), donde 'p' es una variable a sustituir por oraciones y 'x' una variable a sustituir por un nombre de la oración p. Las demostraciones para oraciones μ del tipo: a/b y d son similares.

Trascendencia: $\delta(X) \notin X$.

1. $\delta(X) \in X$, (suposición)
2. $\mu \in X$ (def. $\delta(X)$ + ISVI: 1)
3. $\mu \in T$ ($X \subseteq T$: 2)
4. ***$\forall x\ (\mu = x \rightarrow x \notin X)$*** ***(Esquema-T: 3)***
5. ***$\mu = \mu \rightarrow \mu \notin X$*** ***(Instancia: 4)***
6. ***$\mu = \mu$*** ***(Teorema)***
7. $\mu \notin X$ ***(Modus ponens: 5, 6)*** (Esquema-T: 3)
8. $\delta(X) \notin X$. (absurdo 1-7 por contradicción 2, 7)

Clausura: $\delta(X) \in \Omega$.

1. $\mu \notin X$ ($\delta(X) \notin X$ (trascendencia) + def. δ + ISVI)
2. ***$\mu = t$*** ***(suposición)***
3. ***$t \notin X$*** ***(ISVI: 1, 2)***
4. ***$\mu = t \rightarrow t \notin X$*** ***(Introducción $\rightarrow$: 2-3)***
5. ***$\forall x\ (\mu = x \rightarrow x \notin X)$*** ***(Introducción $\forall$: 4)***
6. $\mu \in T$ ***(Esquema-T: 5)*** (Esquema- T: 1)
7. $\delta(X) \in T$ (def. δ: 6)

Estos razonamientos muestran que la paradoja del mentiroso, incluso bajo presentaciones un tanto complejas que entrañan cuantificación y otras conectivas lógicas, encaja en el Esquema de Inclusión. Además de ofrecer una demostración –las líneas *no* resaltadas en el razonamiento anterior– para el mentiroso simple: $\delta(\Omega) = \mu =$ '$\mu \notin T$', Priest demuestra que también encajan en este esquema la paradoja de

Grelling [8] sobre el término 'heterológico' y las "cadenas de mentirosos" del tipo: μ_1 = 'μ_2 es verdadero'; μ_2 = 'μ_1 es falso'.[22] Según él, estas paradojas se producen en "el límite de lo cognitivo" (son, pues, del tipo [$Y \subset \wp(\Omega)$ y $\Omega \in Y$]). Como ocurre con todas las contradicciones de inclusión, los razonamientos que desembocan en la contradicción pueden aparecer en una prueba por *reductio* destinada a demostrar la imposibilidad de algunos de los elementos necesarios para formular la ejemplificación relevante del esquema EI. Tarski utilizó la paradoja del mentiroso [1] para demostrar que, en un lenguaje que pudiera expresarla, no era posible concebir coherentemente la existencia del conjunto T de todas sus verdades.

Sin embargo, si las paradojas que hemos visto hasta ahora son versiones de la paradoja del mentiroso, entonces, por similares que sean a la paradoja de la cita [11], no pueden coincidir con ella. Como apunta Tarski, esta paradoja es independiente de la presencia en nuestro lenguaje del predicado 'verdadero' o del conjunto T de todas las verdades. ¿Satisface [11] el Esquema de Inclusión sin apelar a la noción de verdad? La respuesta es afirmativa.

Dados un determinado lenguaje, *L*, y una lógica, *l*, en los que sean válidos y expresables todos los principios (en particular: 'P' = 'Q' $\rightarrow$ (P $\leftrightarrow$ Q)) y oraciones que utilizamos para derivar la paradoja de la cita, podemos definir un predicado, 'demostrable', cuya extensión comprenda (sólo) a las oraciones de *L* demostrables de acuerdo con *l*. Consideremos φ = "es una oración de *L* demostrable en *l*" y ψ = "es idéntico a Ω".

[22] Priest 1994a y *BLT*, pp. 143-46.

1) $\Omega = \{y : y$ es una oración de L demostrable en $l\}$ existe y $\Omega = \Omega$.

2) Si $X \subseteq \Omega$ y $X = \Omega$, entonces $\delta(X) = c =$ "$\forall p$ (c = 'p' $\rightarrow \neg p$)".

Trascendencia: $\delta(X) \notin X$.[23]

1 $\delta(X) \in X$	(Suposición)
2 $c \in X$	(definición de δ: 1)
3 c es demostrable en l	($X = \Omega$: 2)
4 $\forall p$ (c = 'p' $\rightarrow \neg p$)	(x es demostrable $\rightarrow p$: 3)
5 c = '$\forall p$ (c = 'p' $\rightarrow \neg p$)' $\rightarrow \neg\forall p$ (c = 'p' $\rightarrow \neg p$)	(Instancia: 4)
6 c = '$\forall p$ (c = 'p' $\rightarrow \neg p$)'	(definición de c)
7 $\neg\forall p$ (c = 'p' $\rightarrow \neg p$)	(Modus Ponens: 5, 6)
8 Contradicción	(4, 7)
9 $\delta(X) \notin X$	(Absurdo: 1-8)

Clausura: $\delta(X) \in \Omega$.[24]

[23] En el paso 4, nos apoyamos en el siguiente principio: "x es demostrable $\rightarrow p$" (donde 'p' se sustituye por una oración y 'x' por un nombre de la oración p). Sabemos que, en general, "x es demostrable $\leftrightarrow p$", no es correcto, en caso contrario, tendríamos mediante el Esquema-T tarskiano: "x es demostrable $\leftrightarrow$ x es verdadera" pero, como sabemos, muchos sistemas formales no son completos, i.e., no todo lo que es verdadero en ellos es demostrable en ellos. Sin embargo, no parece controvertido afirmar que todo lo demostrable sea verdadero. Eso es, a lo sumo, lo que se sigue de "x es demostrable $\rightarrow$ p". En general, "$p \rightarrow x$ es demostrable" no es correcto.

[24] En la segunda parte de la prueba seguimos implícitamente otro principio relativo al predicado 'demostrable': "c *es demostrable si y sólo si existe una cadena de oraciones* $s_0, s_1, s_2,..., s_n$, *tales que cada elemento* s_i $(0 \leq i \leq n)$ *de la cadena o es una premisa o se obtiene a partir de los elementos precedentes según las reglas lógicas pertinentes y* c = 's_n'". Queda claro que este principio está estrechamente ligado a la noción de demostrabilidad y no es equivalente a "$p \rightarrow x$ es demostrable".

1 $\exists p\,(c = `p' \wedge p)$	(Suposición)
2 $c = `q' \wedge q$	(Sup. –prueba por casos, eliminar $\exists$ en 1)
3 $`q' = c = \text{“}\forall p\,(c = `p' \rightarrow \neg p)\text{”}$	(Def. de c + eliminación conjunción: 2)
4 $q \leftrightarrow \forall p\,(c = `p' \rightarrow \neg p)$	($`P' = `Q' \rightarrow (P \leftrightarrow Q)$: 3)
5 q	(Eliminación conjunción: 2)
6 $\forall p\,(c = `p' \rightarrow \neg p)$	(Mod. Pon.: 4 (“$q \rightarrow \forall p\,(\ldots)$”), 5)
7 $\forall p\,(c = `p' \rightarrow \neg p)$	(Eliminación $\exists$: 2-6)
8 c es demostrable en *l*	(def. c + prueba 1-7 con 1 como premisa)
9 $\delta(X) \in \Omega$	(definición de δ y Ω: 8)
10 $\delta(X) \in X$	(definición ψ: 9)
11 Contradicción	(10 + trascendencia: $\delta(X) \notin X$)
12 $\neg\exists p\,(c = `p' \wedge p)$	(Absurdo:1-11)
13 $\forall p\,(c = `p' \rightarrow \neg p)$	(Teorema: 12 $\leftrightarrow$ 13)
14 c es demostrable	(Prueba 1-13)
15 $\delta(X) \in \Omega$	(definición de δ y Ω: 14)

Como se puede apreciar, la presentación bajo el Esquema de Inclusión de la paradoja de la cita sigue los pasos de la demostración especificada en [11]. En este caso, Ω es el conjunto de todas las oraciones de L demostrables según la lógica l y δ se define para uno solo de sus subconjuntos, a saber, el propio Ω. Tenemos, pues, un ejemplo de contradicción de inclusión de la forma [$Y = \{\Omega\}$]. Al igual que ocurre con el resto de contradicciones de inclusión, el razonamiento que aparece en la paradoja de la cita puede ser utilizado para refutar por *reductio* alguno de sus supuestos de partida. Tarski, por ejemplo, utiliza el argumento para rechazar que las comillas tengan un carácter funcional.

De este modo el problema se desvanece ya que, al ser las comillas simples caracteres de un nombre (y no funciones), no podemos cuantificar variables (o, mejor dicho, caracteres) que se encuentren entrecomilladas. Hacerlo sería como ligar con un cuantificador la 'p' de 'carpeta'. Esto hace inviable la posibilidad de interpretar la oración paradójica $\delta(\Omega)$ como requiere la demostración de la contradicción en [11].

2.7 Razonamientos que satisfacen el Esquema de Inclusión: el teorema de Cantor

Todas las "paradojas" que hemos visto hasta ahora [1]-[11] entrañan razonamientos que encajan en el Esquema de Inclusión y que pueden utilizarse (si hemos de rechazar la contradicción que establecen) en reducciones al absurdo que niegan la verdad de alguno(s) de sus supuestos de partida. Es precisamente este rasgo lo que nos ha llevado a hablar preferentemente de "contradicciones de inclusión" (en vez de "paradojas") y a interpretar como tales razonamientos que nunca se han considerado paradójicos al aparecer asociados (desde su origen) a demostraciones por *reductio* ampliamente aceptadas. Éste es el problema hacia el que apuntábamos al inicio de 2.3 y al que volveremos repetidas veces a lo largo de este trabajo. La existencia de paradojas "genuinas", de razonamientos válidos con premisas verdaderas y conclusión falsa, se nos muestra como inaceptable. Por eso insistimos siempre en calificar de aparentes la supuesta validez del razonamiento o la verdad de las premisas. Dentro del Esquema de Inclusión caben razonamientos cuya conversión en una *reductio* no parece sencilla (no está claro qué premisas debemos rechazar en la paradoja del mentiroso, por ejemplo), pero también caben razonamientos cuya conversión en una *reductio* parece incontestable incluso para aquellos que, como Priest, piensan que muchas de estas conversiones no son satisfactorias.[25] Uno de los ejemplos más destacados es el teorema de Cantor, consideraré aquí una de sus formulaciones más generales y mostraré que la *reductio* crucial contiene una contradicción de inclusión, es decir, un razonamiento que encaja en EI.

[I] *Teorema de Cantor.*[26] Dada una función *sobreyectiva* f: A $\rightarrow$ B y el siguiente conjunto: $\Omega = \{x \in \mathrm{A}: x \notin f(x)\}$, se cumple siempre que $\Omega \notin \mathrm{B}$ (Ω no está en el rango de f).

Demostración: Supongamos (con vistas a una *reductio*) que $\Omega \in \mathrm{B}$. Como f es sobreyectiva, *existe* al menos un $a \in \mathrm{A}$ tal que $f(a) = \Omega$. Ahora bien, si $a \in \Omega$, entonces $a \in f(a)$ (puesto que $f(a) = \Omega$), pero en ese caso $a \notin \Omega$ (por la definición de Ω) y tenemos una contradicción. Así pues, $a \notin \Omega$. Sin embargo, si $a \notin \Omega$, entonces $a \notin f(a)$ (ya que $f(a) = \Omega$), de donde concluimos (por la definición de Ω) que $a \in \Omega$. Una nueva contradicción.

[25] Priest (lo veremos en los capítulos 4-5) no acepta la *reductio ad absurdum* de la existencia del conjunto universal que se suele inferir a partir de la paradoja de Russell. Para Priest, incluso el primer teorema de incompletud de Gödel encierra una paradoja genuina que sugiere la inconsistencia de la aritmética.

[26] Tomamos esta versión de Barwise y Moss (1996, p. 26), pero cualquier otra valdría a todos los efectos.

Como no hay más opciones que $a \in \Omega$, o $a \notin \Omega$, rechazamos nuestro supuesto inicial y establecemos $\Omega \notin$ B. □

A partir de este teorema se puede concluir que, dado un conjunto A, el conjunto de sus subconjuntos, $\wp$(A), es siempre mayor que A. Si ambos conjuntos tuviesen el mismo número de elementos, existiría una función biyectiva (una correlación uno a uno), *f*, entre A y $\wp$(A). Al ser biyectiva, f: A $\rightarrow$ $\wp$(A) sería, en particular, sobreyectiva. Ahora bien, esto es imposible porque el conjunto $\Omega = \{x \in$ A: $x \notin f(x)\}$ *es un subconjunto* de A y, por tanto, $\Omega \in \wp$(A), contradiciendo el teorema de Cantor [I].

A continuación mostraremos que la *reductio* que aparece en dicho teorema encaja en EI y es una contradicción de inclusión del tipo [Y = $\{\Omega\}$], igual que la paradoja de la cita [11]. Partamos, como exige la *reductio* en [I], de las siguientes premisas:

(Premisa 1) Dados dos conjuntos A y B, existe una función *sobreyectiva* f: A $\rightarrow$ B.
(Premisa 2) $\Omega = \{x \in$ A: $x \notin f(x)\} \in$ B.

En este caso, la propiedad $\varphi(x)$ será "$x \in$ A $\wedge$ $x \notin f(x)$", la propiedad ψ(X) será "X = Ω" y la función δ: $\{\Omega\} \rightarrow \Omega$ será igual al conjunto $\{(\Omega, a)\}$, donde se cumple que $a \in$ A y $f(a) = \Omega$. La existencia de δ y la corrección de la condición sobre a están aseguradas: Como f es sobreyectiva (Premisa 1) y $\Omega \in$ B (Premisa 2), existe al menos un $x \in$ A tal que $f(x) = \Omega$. Así pues, al definir δ nos hemos limitado a elegir uno de esos x, al que hemos llamado a. Por último, δ: $\{\Omega\} \to \Omega$, es una función de un $Y \subseteq \wp(\Omega)$ a Ω. Es obvio que $\{\Omega\} \subseteq \wp(\Omega)$ y demostraremos que $a \in \Omega$ al demostrar la condición de clausura, ya que $a = \delta(\Omega)$. La ejemplificación pertinente del esquema de inclusión es la siguiente:

1) $\Omega = \{x \in A: x \notin f(x)\}$ existe (y $\Omega = \Omega$).
2) Si $X \subseteq \Omega$ (y $X = \Omega$), entonces $\delta(X) = \delta(\Omega) = a$.

Trascendencia:

1. $\delta(X) \in X$	(Suposición)
2. $\delta(\Omega) \in \Omega$	($X = \Omega$ por la definición de ψ: 1)
3. $\delta(\Omega) \in f(a)$	($\Omega = f(a)$ por la condición sobre a: 2)
4. $a \in f(a)$	($\delta(\Omega) = a$ por la definición de δ: 3)
5. $a \notin \Omega$	(por la definición de Ω: 4)
6. $\delta(\Omega) \notin \Omega$	($\delta(\Omega) = a$ por la definición de δ: 5)
7. Contradicción	(2, 6)
8. $\delta(X) \notin X$	(Absurdo 1-7)

Clausura:

1. $\delta(X) \notin X$	(Suposición)
2. $\delta(\Omega) \notin \Omega$	($X = \Omega$ por la definición de ψ: 1)
3. $\delta(\Omega) \notin f(a)$	($\Omega = f(a)$ por la condición sobre a: 2)
4. $\delta(\Omega) \notin f(\delta(\Omega))$	($\delta(\Omega) = a$ por la definición de δ: 3)
5. $\delta(\Omega) \in \Omega$	($\delta(\Omega) = a + a \in A$ + definición de Ω: 4)
6. Contradicción	(2, 5)
7. $\delta(X) \in X$	(absurdo 1-6)

La contradicción que encontramos en la *reductio* del teorema de Cantor es, pues, otro ejemplo de contradicción de inclusión (de nuevo, un ejemplo del tipo [Y = {Ω}]). En este caso, sin embargo, nadie pone en duda la validez de la *reductio*. Las premisas 1 y 2 sobre las que descansa la inferencia de la contradicción no pueden ser verdaderas conjuntamente, eso es lo que establece el teorema de Cantor. Este teorema es, de hecho, clave en la propia génesis de la teoría de conjuntos ya que prueba, entre otras cosas, que no todos los conjuntos infinitos tienen la misma cardinalidad (si A es un conjunto infinito, entonces $\wp$(A) es un conjunto infinito mayor). La contradicción que encierra la demostración del teorema no es considerada hoy en día como una paradoja por nadie. Es importante recalcar este hecho para no perder de vista que lo que el esquema capta no es aquello que unifica a todas las paradojas, sino aquello que comparten ciertos razonamientos (paradójicos o no) a los que hemos denominado contradicciones de inclusión. Hecha esta puntualización, cabe señalar que el Esquema de Inclusión cubre un amplio espectro de razonamientos y que un gran número de paradojas encaja en sus parámetros. En el apéndice (2.10) de este capítulo mostramos algunas de las paradojas que encajan en EI. Sólo nos detendremos en aquellos ejemplos que Priest no aborda explícitamente (salvo la paradoja de Yablo [13] que incluimos por su singularidad), las demás o no son explícitamente reducidas al Esquema de Inclusión por Priest (como la paradoja de Curry [17]), o no aparecen ni en *BLT* ni en otros escritos de Priest (como la paradoja de Barwise y Moss [12] o la del hiper-juego [15]).

2.8 Críticas al Esquema de Inclusión

Para finalizar este capítulo consideraremos una serie de críticas planteadas al Esquema de Inclusión por diferentes comentaristas.

2.8.1 Las paradojas de Yablo [13] y Curry [17] no encajan en EI

Comencemos por la paradoja de Yablo [13]. Priest (1997a) expresa esta paradoja en términos de EI (véase 2.10.2 en el apéndice de este capítulo). Existen dudas, sin embargo, con respecto a si lo que Priest expresa mediante EI es la paradoja de Yablo u otra distinta relacionada de algún modo con ésta. Sea como fuere, debería quedar claro que la paradoja de Yablo no puede entenderse en ningún caso como un contraejemplo al Esquema de Inclusión. Si Priest (1997a) tiene razón y la única manera de entender la paradoja es reformulándola en el sentido propuesto por él (véase 2.10.2), entonces la paradoja de Yablo encaja en EI. Si se equivoca, por contra, y la paradoja de Yablo es fundamentalmente distinta de la que él reformula, entonces lo único que importa es ver si su variante de la "paradoja de Yablo" encaja en EI, porque es la única en la que la autorreferencia o la circularidad juegan un papel crucial. Si la paradoja de Yablo no fuese reducible a la interpretación de Priest, tendríamos una prueba clara de que –como afirma el propio Yablo (1993)– *no es una paradoja reflexiva*. Pero entonces *no tiene por qué encajar en EI*, simplemente pertenece a otra familia de paradojas. El Esquema de Inclusión no pretende ofrecer la estructura de *todas* las paradojas, sino sólo la de aquellas que Russell denominó "paradojas reflexivas" y que involucran algún tipo de circularidad. Muchos conciben la paradoja de Yablo como un ejemplo de paradoja no circular y han desarrollado versiones de la misma que no involucran oraciones sino pensamientos, conjuntos, etcétera.[27] Si están en lo cierto, la paradoja no pertenece a la familia de las contradicciones de inclusión.

Un problema más grave es el que supone la paradoja de Curry [17]. Si interpretamos '$\rightarrow$' como un condicional material, la paradoja de Curry encaja en EI (véase 2.10.6 en el apéndice de este capítulo). Sin embargo, Priest admite que si el condicional se entiende de otra manera, entonces la paradoja de Curry deja de encajar en EI. Según Priest (*BLT*, pp. 168-69, 278; también en Priest 1994a), en una lectura clásica del implicador '$\rightarrow$' como condicional material, las oraciones '$\alpha \rightarrow \perp$' y '$\neg\alpha$' son equivalentes, cosa que no ocurre cuando interpretamos '$\rightarrow$' de otra manera. Pensemos en la paradoja de Curry: s = 's es verdadera $\rightarrow \perp$'. Si '$\rightarrow$' es un condicional material, tenemos que: (s es verdadera) $\leftrightarrow$ (s es verdadera $\rightarrow \perp$) $\leftrightarrow \neg$(s es verdadera). El primer bicondicional es una instancia del Esquema-T, el segundo se cumple por ser '$\rightarrow$' un condicional material. Uniendo ambos hechos obtenemos: (s es verdadera)

[27] Sorensen 1998 considera una fila infinita de personas en la que cada una piensa de sus sucesores en la fila que "lo que piensan es falso". Goldstein 1994 da una versión de la paradoja de Yablo para conjuntos: En este caso se considera una secuencia de conjuntos C_n (uno por cada número natural) en la que cada conjunto C_n satisface la condición $\forall x$ ($x \in C_n \leftrightarrow \forall k > n, x \notin C_k$.). Sobre esta paradoja puede verse también Beall 1998, Tennant 1995 y Hardy 1995.

$\leftrightarrow \neg$(*s* es verdadera), así pues, *s* es, en el fondo, una variante de la paradoja del mentiroso que, como ya hemos visto en 2.6, encaja en el Esquema de Inclusión. El problema que presenta esta paradoja es que, aun cuando '$\rightarrow$' no fuese un condicional material –y se rompiera por tanto la equivalencia (*s* es verdadera) $\leftrightarrow \neg$(*s* es verdadera)–, *s seguiría siendo una paradoja* en todos aquellos sistemas lógicos que aceptasen el *Principio de Absorción*: "$\alpha \rightarrow (\alpha \rightarrow \beta) \Rightarrow \alpha \rightarrow \beta$". El razonamiento es el siguiente:

1. (*s* es verdadera) $\leftrightarrow$ (*s* es verdadera $\rightarrow \perp$)	(Esquema-T)
2. *s* es verdadera $\rightarrow \perp$	(Principio de Absorción: 1 ($\rightarrow$))
3. *s* es verdadera	(Esquema-T: 2)
4. $\perp$	(Modus Ponens: 2, 3)

Como se puede apreciar, el absurdo aquí depende sólo de reglas relacionadas con la implicación. Ni la negación ni, por consiguiente, la paradoja del mentiroso juegan un papel relevante.

T. Williamson 1996 (p. 333), P. Grim 1998 (p. 721) e I. Grattan-Guinness 1998 (pp. 826-27) se aferran a estas observaciones de Priest para concluir, contra él, que el Esquema de Inclusión no cubre todas las paradojas reflexivas y que, por ello, no proporciona una caracterización satisfactoria de las paradojas que pretende describir. Quizá la paradoja de Yablo se pueda ignorar a la hora de ofrecer una descripción unificada de las paradojas "reflexivas" o de "circularidad", pero la de Curry no, porque cualquiera de sus formulaciones es un ejemplo manifiesto de autorreferencia.

La respuesta que Priest ofrece a este problema no parece convencer a sus críticos. Pese a ello resulta plausible. Según Priest, todo lo que se le puede pedir al Esquema de Inclusión es que dé cuenta de la paradoja de Curry cuando la *negación* juega un papel relevante en ella. La negación ha de hallarse forzosamente presente en todas las *contradicciones* de inclusión: sin negación, no puede haber contradicción. Si la paradoja de Curry no descansa en ningún sentido sobre reglas relacionadas con la negación ni precisa ningún tipo de negación en su formulación, entonces no puede ser una *contradicción* de inclusión.[28] Priest parece estar en lo cierto al afirmar que la paradoja de Curry (sin condicional material *ni negación*) pertenece a una familia distinta de paradojas. El único problema que plantea su respuesta es su insistencia previa en que *las paradojas de circularidad* (o *autorreferencia*) satisfacen el Esquema de Inclusión. Si éste es todo el problema, entonces bastaría con abandonar la expresión paradojas "reflexivas" al hablar de contradicciones de inclusión. Pero, frente a la evidencia aportada por Priest, sería absurdo negar que existe una familia de paradojas unidas por rasgos relevantes y no arbitrarios que comparten una misma estructura: el Esquema de Inclusión.

En realidad, el único principio relativo a la negación que necesitamos para reducir la paradoja de Curry a EI es el de *reductio ad*

[28] "In this case, the curried versions of the paradoxes belong to a quite different family. Such paradoxes do not involve negation and, *a fortiori*, contradiction. They have therefore nothing to do with contradictions at the limits of thought." *BLT*, p. 169.

absurdum (como se puede apreciar en la prueba de *trascendencia* de 2.10.6).[29] Cualquier concepción del condicional que sea compatible con la validez de las demostraciones por *reductio*, permite formular la paradoja de Curry dentro de los parámetros de EI. La paradoja de Curry sólo escapa a EI al erradicar por completo la negación (y todas las reglas de inferencia relacionadas con ella) de su formulación. Pero cuando esto ocurre, parece claro que hablamos de otro tipo de paradoja.

2.8.2 ¿Define EI las paradojas en términos de condiciones necesarias y suficientes?

Grattan-Guinness 1998 (p. 831) critica a Priest por el hecho de que EI no ofrece condiciones necesarias y suficientes para identificar una paradoja. No ofrece condiciones necesarias porque hay paradojas –Curry, sin negación– que no encajan en el esquema. No ofrece condiciones suficientes porque hay razonamientos claramente falsos –la paradoja del barbero [18] (ibíd., pp. 827-28)–,[30] o claramente coherentes –el primer

[29] Si este principio es válido, es muy fácil que se cumpla la equivalencia "$(\alpha \rightarrow \perp) \leftrightarrow \neg\alpha$". Para ello sólo necesitaríamos que nuestro condicional respetase las leyes de *reductio* y *modus ponens*. También necesitaríamos que toda contradicción se considerase un absurdo: $\perp \leftrightarrow (\alpha \wedge \neg\alpha)$:
[Demostración de $(\alpha \rightarrow \perp) \rightarrow \neg\alpha$]: 1. Supóngase $\alpha \rightarrow \perp$; 2. Supóngase α; 3. $\perp$ (*modus ponens*: 1, 2); 4. $\neg\alpha$ (*reductio*: 1-3); 5. $(\alpha \rightarrow \perp) \rightarrow \neg\alpha$ (Introducción '$\rightarrow$' 1-4).
[Demostración de $\neg\alpha \rightarrow (\alpha \rightarrow \perp)$]: 1. Supóngase $\neg\alpha$; 2. Supóngase α; 3 $\perp$ ($\perp \leftrightarrow (\alpha \wedge \neg\alpha)$: 1 + 2); 4. $\alpha \rightarrow \perp$ (Introducción '$\rightarrow$': 2-3); 5. $\neg\alpha \rightarrow (\alpha \rightarrow \perp)$ (Introducción '$\rightarrow$': 1-4).
Esto justifica, en parte, la afirmación de Priest según la cual la paradoja de Curry encaja en EI cuando se cumple "$(\alpha \rightarrow \perp) \leftrightarrow \neg\alpha$". Aunque, siendo estrictos, sólo necesitamos la ley de *reductio ad absurdum*.

[30] [18] La paradoja del barbero. Imaginemos un barbero que afeita sólo a aquellas personas que no se afeitan a sí mismas. Ese barbero se afeita a sí mismo sí y sólo sí no se afeita a sí mismo. En *BLT* (p. 157, nota 2), se muestra que esta paradoja encaja en EI y es de tipo $[Y \subseteq \wp(\Omega) \wedge \Omega \in Y]$: Sea $\varphi(y)$ = '*y* es una persona' y $\psi(X)$ = 'X es un

teorema de incompletud de Gödel (ibíd., p. 831)– que encajan en el esquema sin ser paradójicos.

Ya hemos dicho que Priest no pretende ofrecer condiciones necesarias y suficientes para identificar *todas* las paradojas sino sólo para identificar una familia de paradojas con rasgos característicos. Hemos admitido también que esa familia no debe incluir paradojas que prescindan de la negación y que, habida cuenta de ciertas versiones de la paradoja de Curry, es preferible hablar de contradicciones de inclusión (o circulares) antes que de paradojas reflexivas. Si nos centramos en las contradicciones de inclusión paradójicas, EI ciertamente ofrece condiciones necesarias. El problema ahora es, ¿debe EI proporcionar condiciones suficientes? La respuesta es "no". EI ofrece la estructura formal que comparten ciertos razonamientos: las contradicciones de inclusión. Que en un contexto determinado una contradicción de inclusión se considere verdadera (y, por tanto, genuina o irremediablemente paradójica) o no depende de cual sea el estatus de sus premisas. Las paradojas sólo aparecen cuando nos resulta imposible rechazar alguna de las premisas de EI (cualquier elemento o regla de inferencia necesarios para derivar $\delta(\Omega)$ o para construir Ω, φ, ψ o δ bajo las condiciones estipuladas por EI). Cuando alguna de estas premisas es falsa, podemos rechazar la verdad de la conclusión, i.e., de la

---

grupo de personas en el que aquellos que no se afeitan a sí mismos son afeitados por una sola persona'. Ω = {y: y es una persona}, $\delta(X)$ es la persona que *sólo* afeita a los miembros de X que no se afeitan a sí mismos. La condición de clausura se sigue de inmediato ($\delta(X)$ es una persona) y la de trascendencia por *reductio* (si $\delta(X) \in X$, entonces se afeita y no se afeita a sí mismo, por tanto $\delta(X) \notin X$). Según Priest, ésta no es una paradoja genuina porque $\psi(\Omega)$ es claramente falso.

contradicción. Éste es el caso de la paradoja del barbero (donde rechazamos $\psi(\Omega)$, véase la nota anterior) y también del teorema de Gödel, o de cualquier otro enunciado que se presente como teorema pero que incluya una contradicción de inclusión. EI sólo ofrece condiciones necesarias y suficientes para determinar contradicciones de inclusión. Como hemos dicho ya, que haya paradojas irresolubles o no es un hecho altamente controvertido, por otra parte, que algo sea considerado como una paradoja depende de factores cambiantes relativos a la evolución de nuestras convicciones sobre la verdad de ciertos enunciados y la validez de ciertas inferencias. Lo que antaño se consideró una paradoja, es hoy a menudo parte de una prueba por *reductio* destinada a mostrar la falsedad de alguna de las premisas del razonamiento paradójico. Lo que aquí buscamos es la forma que tiene la inferencia de esta contradicción, tanto si podemos ubicarla en una *reductio* como si no. Ésta era la razón de ser de la sección 2.7 de este trabajo.

2.8.3 Algunos elementos de EI parecen redundantes en algunas paradojas

Tennant (1998, pp. 29-30) defiende que algunos de los elementos estructurales que aparecen en el Esquema de Inclusión son irrelevantes a la hora de expresar ciertas paradojas. La paradoja del mentiroso, por ejemplo, no requiere ni en su formulación ni en la derivación de una contradicción que mencionemos al conjunto T de todas las verdades, como hace Priest.

Una de las razonas por las que Priest (y también Russell) apelan al conjunto T de todas las verdades es porque formulan la paradoja en términos de conjuntos. Así pues 'x es verdadero' se representa por medio de '$x \in T$', igual que 'x es un caballo' se representaría por medio de '$x \in C$', siendo C el conjunto de todos los caballos. Tennant también critica este tipo de traducciones a lenguaje conjuntista de todas las paradojas, pero creo que estas quejas son injustas.

Decir que las paradojas tienen algo en común y que el Esquema de Inclusión capta ese algo que todas ellas comparten, no es decir que todas sean iguales. Las paradojas son diferentes, los conceptos que se utilizan son diferentes y las diferencias se han de tener en cuenta. Sin embargo, también encontramos similitudes y la única manera de expresar estas similitudes de forma asequible es buscar un esquema estructural de carácter general en el que quepan todas. Si elegimos expresar ese esquema, por razones prácticas, en términos conjuntistas, entonces tendremos que "traducir" a ese "lenguaje" algunas paradojas que no hacen referencia explícita a conjuntos. Pero también sería posible ofrecer una versión del esquema en términos de propiedades, prescindiendo de conjuntos.[31] Por otra parte, si el esquema es general, habrá, por supuesto, casos límite en los que ciertos ingredientes de la paradoja no jugarán un papel determinante, como ocurre con ψ en el caso de las contradicciones de inclusión que encajan en el Esquema de Russell *simpliciter*. Sin embargo, todos estos rasgos divergentes son los que nos permiten

[31] Priest lleva a cabo esta reformulación en *BLT*, p. 280, nota 26 y discute (pp. 279-80) el estrecho vínculo existente entre "satisfacer una propiedad" y "pertenecer a un conjunto".

encontrar subfamilias dentro de una misma familia de paradojas. Lo que hace que EI sea valioso es que nos permite describir de forma unificada un buen número de paradojas sin ejercer violencia sobre ninguna de ellas, es decir, sin que el encaje de una nueva paradoja introduzca elementos en EI que nos fuercen a desalojar del esquema una paradoja que antes encajaba en él o uno de los rasgos descriptivos que articulan EI. Todos los elementos de EI tienen una interpretación natural dentro de cada una de sus ejemplificaciones.

2.8.4 La clasificación de Ramsey subsiste

Tennant afirma (1998, pp. 28-30) que Priest no consigue desacreditar la distinción de Ramsey. Sus argumentos se basan, en parte, en la artificiosidad del encaje de la paradoja del mentiroso (ya hemos tratado esta cuestión). Cabe señalar ahora (como hicimos en 2.5) que la distinción que Ramsey introduce entre paradojas conjuntistas y semánticas subsiste de algún modo. Sin embargo, no es menos cierto que Priest consigue dos cosas: en primer lugar, muestra que existe un patrón estructural básico que comparten *todas* las paradojas reflexivas (con negación), semánticas o conjuntistas; en segundo lugar, muestra que, apoyándonos en el Esquema de Inclusión, es posible establecer *sub*divisiones con mucho más rigor del aportado por Ramsey. Distinciones que se fundan en aspectos estructurales relativos al modo en que diferentes paradojas satisfacen EI. Desde este punto de vista, Priest tiene éxito (frente a Ramsey) al ofrecer un marco estructural en el que encajan las paradojas reflexivas. Pero, al mismo tiempo, EI nos permite formular de una forma más precisa, apelando al peso específico que cobran en cada paradoja ciertos rasgos estructurales, en qué sentido las circunstancias de cada contradicción de inclusión son relevantes.

2.8.5 Con algún artificio, toda contradicción encaja en EI

Priest reconoce (*BLT*, pp. 135-36) que, si admitimos como casos límite del esquema aquellos en los que la propiedad ψ es "ser igual a Ω", entonces, con algún artificio, cualquier contradicción, $p \wedge \neg p$, podría pasar por una contradicción de inclusión. En estos casos tendríamos: φ es "$(x = p) \wedge (x \wedge \neg x)$", ψ es "$X = \Omega$", $\Omega = \{p \wedge \neg p\}$ y $\delta(X) = p$. Priest responde a su propia objeción diciendo que mediante artificios, cualquier caso límite trivial puede encajar en un esquema general. La virtud de un esquema radica en que sea capaz de señalar y articular los elementos esenciales que se dan cita en los casos *no triviales*, los casos triviales o límite son hasta cierto punto inevitables pero no son ninguna novedad ni ningún impedimento cuando emprendemos la tarea de ofrecer una clasificación sistemática de objetos en cierto campo. Después de todo, el conjunto vacío se define como un conjunto, el cero es un número (aunque ni los romanos ni los griegos pensaron tal cosa). A veces, definimos las constantes de un lenguaje como "funciones" que asignan el mismo valor (constante) a cada argumento; los predicados se definen como "relaciones monádicas", etcétera. Si las contradicciones simples no encajasen en EI, seguramente no podríamos demostrar contradicciones a partir de él. Por otra parte, no podemos excluir del esquema el caso en el que ψ es la propiedad "X = Ω", de lo contrario algunas contradicciones de inclusión muy importantes no encajarían en EI.

2.9 Conclusiones

El Esquema de Inclusión nos ofrece una descripción estructural unificada de una serie de razonamientos en los que encontramos elementos constantes o recurrentes que desembocan en una contradicción. En este esquema encontramos siempre una totalidad A y una operación f cuyas definiciones son interdependientes: A se puede concebir como el resultado de aplicar la operación f sobre los subconjuntos de A. La circularidad de la situación radica en el hecho de que A no se puede definir sin definir antes f(A) que, a su vez, no se puede definir sin determinar A. El papel esencial que juega la negación en estas circunstancias se pone de manifiesto cuando nos percatamos de que, por definición, f, es siempre una operación cuyo resultado satisface la condición: $f(x) \notin x$.

2.10. Contradicciones de Inclusión. (Apéndice capítulo 2)

En Priest 1994a y *BLT* podemos encontrar reducciones a EI de las paradojas [1]-[10]. En *BLT* se reducen a dicho esquema las paradojas [18] y [19] y se apuntan también reducciones de [16] y [17]. En el presente trabajo hemos mostrado que [11] y el teorema de Cantor caben en EI como casos límite (paradojas en las que ψ es la propiedad "ser idéntico a Ω"). A continuación mostraremos cómo encajan en EI las paradojas: [12], [14] y [15]. Incluiremos una reducción explícita de [17] a

EI. Por su peculiaridad, también reproduciremos la reducción de [13] a EI, que se puede encontrar en Priest 1997a.

2.10.1 La paradoja de la denotación de Barwise y Moss [12]

[12] Partiendo de un lenguaje, *L*, cuyos términos denotativos sólo pueden tener *un* referente en cada contexto de evaluación, Barwise y Moss[32] consideran el término denotativo, τ, cuya interpretación se define circularmente del siguiente modo. Dado un contexto de evaluación:

I. τ designa a 1, si τ designa a 0.

II. τ designa a 0, si τ designa a 1.

III. τ designa a 1, si τ carece de referencia.

Si τ designa a 1, entonces designa a 0, lo cual es imposible. Si τ designa a 0, entonces designa a 1, lo que también es imposible. Por tanto τ no puede tener referencia, ahora bien, si ése es el caso, τ designa a 1. Todos los supuestos conducen a una contradicción.

Lo esencial en esta paradoja es considerar un término τ cuya interpretación pueda describirse del siguiente modo: (τ_1) "si τ designa a *a*, entonces no designa a *a*, y si τ carece de referencia, entonces designa a *a*" o, más sencillo aún, (τ_2) "si τ designa a *a*, entonces $a \neq a$, y si τ carece de referencia, entonces designa a *a*". Cualquier término interpretado según (τ_1) o (τ_2) es problemático en el sentido apuntado por Barwise y Moss. Como veremos enseguida, esta paradoja también encaja en el Esquema de Inclusión, siendo del tipo [$Y \subset \wp(\Omega)$ y $\Omega \in Y$]. Para

[32] En Barwise y Moss 1996, pp. 57-8.

demostrarlo, consideremos un lenguaje L interpretado mediante un modelo $<D, I>$, donde D es un universo de interpretación no vacío que contiene al menos el objeto a (i.e. $a \in D$); I es la función que interpreta los términos no lógicos de L a partir de D. En esta ocasión, tomaremos un término, τ, interpretado según (τ_2). La propiedad φ será igual a "... es una expresión denotativa con referencia", formalmente: "$\exists y\ (y \in D \wedge I(...) = y)$"; y la propiedad ψ, igual a "... es definible". (Ω será, por tanto, el conjunto de todas los términos de L con un referente en D según I):

1) $\Omega = \{\alpha \in L: \exists y\ (y \in D \wedge I(\alpha) = y)\}$ existe y Ω es definible.
2) Si $X \subseteq \Omega$ y X es definible, $\delta(X) = \tau$ (donde $\tau \in L$).

Definición de $I(\tau)$: (1) $\tau \in X \rightarrow I(\tau) \neq I(\tau)$.
(2) $\tau \notin X \rightarrow I(\tau) = a$ (donde $a \in D$).[33]

Trascendencia: Supongamos que $\delta(X) \in X$, entonces $\delta(X) \in \Omega$ (ya que $X \subseteq \Omega$), o, lo que es igual, $\tau \in \Omega$ (ya que $\delta(X) = \tau$). Ahora bien, dadas las definiciones de Ω e $I(\tau)$, esto significa que existe un $y \in D$ tal

[33] Como δ ha de ser una función, hemos de asegurar que para todo subconjunto definible X de Ω, exista un *único* término τ, tal que $\delta(X) = \tau$. Esto se puede conseguir de dos maneras: (A) O bien consideramos un único término τ en L cuya interpretación se defina del modo especificado; (B) o bien consideramos que L contiene un término τ para cada X *definible* incluido en Ω (no puede haber uno para todo $X \subseteq \Omega$ sin restricciones, de lo contrario la cardinalidad de Ω y la de $\wp(\Omega)$ coincidirían, lo que es imposible por el teorema de Cantor [I]). En (A), δ siempre asigna el *mismo* término τ a cada $X \subseteq \Omega$ (definible), sin embargo la *interpretación* de τ varía, siendo sólo paradójica si $X = \Omega$. En (B), δ asigna un τ diferente, pero único, a cada X definible, $I(\tau)$ es, en cada caso, el

que $I(\tau) = y \wedge y \neq y$. Como no puede existir ningún y de estas características, rechazamos el supuesto inicial y obtenemos $\delta(X) \notin X$ (i.e., $\tau \notin X$).

Clausura: sabemos (por "trascendencia") que $\tau \notin X$, a partir de la definición de $I(\tau)$, podemos inferir que $I(\tau) = a$ (donde $a \in D$), por lo tanto, $\exists y\ (y \in D \wedge I(\tau) = y)$. Concluimos, pues, que $\tau = \delta(X)$ tiene referencia, i.e., $\delta(X) \in \Omega$.

Según L contenga un τ para cada $X \subseteq \Omega$ definible, o uno sólo para todos (véase la nota anterior), podemos utilizar la paradoja que se sigue de $\delta(\Omega)$ para demostrar por *reductio* que no puede existir $\tau = \delta(\Omega)$ o, alternativamente, que no podemos definir $I(\tau)$ si Ω está en el dominio de δ. Otros elementos de esta ejemplificación de EI también se podrían poner en duda (la existencia de δ, por ejemplo).

2.10.2 La paradoja de Yablo [13]

[13] La paradoja de Yablo se formula mediante una serie infinita de oraciones s_0, s_1, s_2, ... en la que cada oración de la serie dice de sus sucesoras que son falsas:

(s_0) Para todo $k > 0$, s_k no es verdadera.

(s_1) Para todo $k > 1$, s_k no es verdadera.

(s_2) Para todo $k > 2$, s_k no es verdadera. ...

---

resultado de reemplazar (en la anterior descripción) X por su definición. En (B) sólo un τ es paradójico: $\tau = \delta(\Omega)$.

Si simbolizamos el predicado 'verdadero' mediante 'T', podemos representar formalmente cualquier oración, s_n, de la serie del siguiente modo: s_n = '$\forall k > n, \neg T(s_k)$'. Así pues, supongamos (para cualquier n) que s_n es verdadera, i.e. $T(s_n)$:[34]

1. (*) $T(s_n)$	$\Rightarrow$	$\forall k > n, \neg T(s_k)$	1'. (*) $T(s_n)$	$\Rightarrow$	$\forall k > n, \neg T(s_k)$
2.	$\Rightarrow$	$\neg T(s_{n+1})$	2'.	$\Rightarrow$	$\forall k > n+1, \neg T(s_k)$
			3'. (*)	$\Rightarrow$	$T(s_{n+1})$

Como se puede apreciar mediante 1-2 y 1'-3', es posible llegar a una contradicción a partir de $T(s_n)$, por lo tanto hemos de concluir: $\neg T(s_n)$. Al ser n arbitrario, es posible inferir por cuantificación universal: $\forall k$, $\neg T(s_k)$, de donde se sigue, en particular: $\forall k > 0, \neg T(s_k)$. Dado que '$\forall k > 0, \neg T(s_k)$' = s_0, tenemos $T(s_0)$ lo que entraña una contradicción, pues habíamos demostrado que $\forall k, \neg T(s_k)$.

Yablo (1993) mantiene que en esta paradoja no hay indicios de circularidad o autorreferencia. Ninguna oración habla de sí misma o de sus predecesoras en la serie, por lo tanto Russell se equivocó al pensar que era condición necesaria de toda paradoja que involucre el término 'verdadero' la presencia de algún tipo de autorreferencia o circularidad. Priest (1997a) ha argumentado de forma ingeniosa que, en la base de la paradoja de Yablo, podemos identificar una estructura circular o autorreferente que, además, encaja en el Esquema de Inclusión.

[34] Presentamos el razonamiento de la paradoja tal y como lo hace Priest (1997a) para facilitar la discusión ulterior.

Volvamos al argumento que establece una contradicción a partir de las oraciones s_0, s_1, s_2, ... Priest nos pide que prestemos atención a la justificación de las líneas marcadas con (*): 1, 1' y 3'. Lo más natural es asumir que se infieren a partir del Esquema-T tarskiano pero esto es un error. El Esquema-T *sólo opera sobre oraciones*, sin embargo, *ni s_n ni s_{n+1} son oraciones*. Al tener una variable libre (a saber, n), las fórmulas '$\forall k > n, \neg T(s_k)$' y '$\forall k > n+1, \neg T(s_k)$' no pueden considerarse oraciones, sólo pueden ser *predicados*. Ahora bien, si hablamos de predicados y no de oraciones, parece más apropiado también hablar de "satisfacción" que de "verdad". Según Priest, una vez nos hemos percatado de este hecho, debemos reformular por completo la paradoja. En la nueva versión partimos de la relación de satisfacción (para predicados monádicos): $S(x, y)$ (x satisface a y), donde x es un objeto e y un predicado; y del predicado $\underline{s}$ = '$\forall k > x, \neg S(k, \underline{s})$'. En vez del Esquema-T utilizamos ahora (para predicados monádicos) el esquema: $S(x, y) \leftrightarrow y(x)$, según el cual x satisface a y si y sólo si x es (un) y. Para todo número natural n, s_n es el nombre de la oración '$\forall k > n, \neg S(k, \underline{s})$', es decir, el resultado de sustituir en $\underline{s}$, la variable x por el número natural n. La paradoja aparece ahora como un caso claro de circularidad ya que "involucra un predicado, $\underline{s}$, de la forma $\forall k > x, \neg S(k, \underline{s})$. El hecho de que $\underline{s}$ = '$\forall k > x, \neg S(k, \underline{s})$' muestra que tenemos aquí un punto fijo, $\underline{s}$, con exactamente el mismo tipo de autorreferencia que asociamos al punto fijo del mentiroso [μ = '$\neg T(\mu)$']. En resumen, $\underline{s}$ es el predicado 'ningún número mayor que x satisface este predicado'".[35]

[35] "...the paradox concerns a predicate, $\underline{s}$, of the form $\forall k > x, \neg S(k, \underline{s})$; and the fact that $\underline{s}$

Bajo esta presentación, la paradoja de Yablo encaja en el Esquema de Inclusión y pertenece a la misma subfamilia que la paradoja del mentiroso. Para ello consideremos φ = "*n* satisface *p*", donde *n* es una variable a sustituir por un número natural y *p* es un predicado monádico (aritmético); como de costumbre, ψ = "es definible".

1) $\Omega = \{(n, p)\text{: } \mathrm{S}(n, p)\}$ existe y Ω es definible.

2) Si $\mathrm{X} \subseteq \Omega$ y X es definible, $\delta(\mathrm{X}) = (0, r_{\mathrm{X}})$

'r_{X}' es el predicado '$\underline{\mathrm{s}} \wedge \forall k > 0, (k, \underline{s}) \notin \mathrm{X}$' ($\underline{\mathrm{X}}$ es un nombre de X y $\underline{s}$, como antes, de $\forall k > x, \neg\mathrm{S}(k, \underline{s})$). Priest demuestra que δ satisface las condiciones de trascendencia y clausura.[36]

Si la única formulación válida de la paradoja de Yablo es la que Priest apunta, entonces la paradoja de Yablo encaja en EI. Pero, ¿es esto cierto? Priest reconoce que la primera versión del argumento que lleva a la contradicción se puede dejar como está -es decir, sin las modificaciones que él sugiere- si en vez de interpretar *n* como una variable libre, la interpretamos como una letra esquemática que representa a cualquier número natural. En este caso, el anterior razonamiento ofrece un número infinito de pruebas de $\neg\mathrm{T}(s_n)$ (una por cada número natural *n*). La inferencia de $\forall k, \neg\mathrm{T}(s_k)$ se obtiene ahora aplicando una (ω-)regla infinitaria del tipo "$\alpha(0), \alpha(1), \alpha(2), \ldots \Rightarrow \forall k$,

= '$\forall k > x, \neg\mathrm{S}(k, \underline{s})$' shows that we have a fixed point, $\underline{s}$, here, of exactly the same self-referential kind as in the liar paradox. [...] $\underline{s}$ is the predicate 'no number greater than x satisfies this predicate'." Priest 1997a, p. 238.

[36] Ibíd., pp. 241-42.

$\alpha(k)$", en la que se infiere algo a partir de un número infinito de premisas. Priest ofrece una serie de objeciones a esta interpretación basadas en nuestras limitaciones epistémicas y nuestra incapacidad para aplicar ω-reglas, pero, en última instancia, admite que un ser como Dios podría aplicar estas reglas. Por otra parte, la regla podría existir y ser válida con independencia de que los seres humanos pudiésemos aplicarla o no. Sin embargo Priest insiste en una idea que parece relevante. Aunque existiesen reglas infinitarias, hay algo en la *formulación* de la paradoja de Yablo que sugiere la presencia de algún tipo de circularidad o autorreferencia. Los *nombres* 's_0', 's_1', 's_2', etc. que designan a las oraciones de la serie tienen un carácter funcional. En ellos, s es en el fondo una función tal que $s(0)$ = '$\forall k > 0, \neg T(s(k))$', etcétera. De hecho, nos resulta posible expresar la paradoja porque existe esta función. *Sólo* así podemos decir que consta de una serie infinita de oraciones generada por una función s (del conjunto de los números naturales al de las oraciones) tal que, para todo n, la oración $s(n)$ es '$\forall k > n, \neg T(s(k))$'. La circularidad radica en que el argumento de la función s aparece *forzosamente* en el valor que s asigna a dicho argumento: sólo podemos formular la oración '$\forall k > n, \neg T(s(k))$' apelando a n, el número que forma parte de su nombre, '$s(n)$'; conversamente, sólo podemos construir el nombre, '$s(n)$', de la oración '$\forall k > n, \neg T(s(k))$', apelando a un rasgo formal de dicha oración, la presencia en ella del número n.

2.10.3 La paradoja del cocodrilo [14]

Un niño cae a las aguas del Nilo y es capturado por un cocodrilo. El padre suplica al cocodrilo que le devuelva a su hijo, pero éste le promete devolvérselo sólo si adivina qué va a hacer con él y comérselo en caso contrario. Tras pensar un momento, el padre responde: "Te lo comerás", con lo que el cocodrilo no puede cumplir su promesa en ningún caso.

Si el padre hubiese respondido "No te lo comerás", el cocodrilo podría haber cumplido su promesa en ambos casos, i.e., devolviendo al niño (ya que el padre habría acertado) o comiéndoselo (ya que el padre no habría acertado). En este sentido la paradoja del cocodrilo se parece a enunciados que *sólo en determinados contextos* desembocan en paradojas, como ocurre con el enunciado (atribuido a Epiménides el cretense): "Todos los cretenses mienten siempre". Si nos encontramos en un contexto en el que se cumple contingentemente que todos los cretenses han mentido siempre hasta el momento de emitir este enunciado y en el que, además, el enunciado es emitido por un cretense, tenemos una paradoja (Epiménides miente si, y sólo si, no miente).

La paradoja del cocodrilo se parece a la de Epiménides también en otro aspecto. Ambas pueden presentarse como paradojas que involucran la noción de verdad al hablar de "mentir" en un caso y de "adivinar" o "prometer" en otro. Olvidando detalles irrelevantes podemos describir el contexto o situación abstracta que plantea la paradoja del cocodrilo del siguiente modo. Se cumplen dos principios (que podemos interpretar como promesas, predicciones, reglas o de otro modo): (I) y (II), y un hecho contingente: (III).

(I) Si x dice "p" y "p" es verdadero, entonces "$\neg q$" es verdadero.
(II) Si x dice "p" y "p" no es verdadero, entonces "q" es verdadero.
(III) x dice "q".

La paradoja surge cuando x dice "q" porque entonces tenemos que el enunciado "q" es verdadero si, y sólo si, "$\neg q$" es verdadero.

Una vez expresada de esta forma, la paradoja no deja de ser una versión contingente o contextual del mentiroso. Lo único que cambia es la manera que tenemos de construir un enunciado, q, tal que $q \leftrightarrow \neg q$. En este caso, la posibilidad de construir dicho enunciado depende del cumplimiento en un contexto de condiciones y hechos independientes de nuestra definición de "verdad" y de la referencia de los términos que aparecen en la oración problemática.[37] Aquí debemos tener en cuenta, además, la presencia *contingente* de principios como (I) y (II) (que podemos interpretar como promesas, predicciones, reglas, etc.) y de hechos como (III). Sin alguno de estos principios o hechos *no se puede derivar una contradicción a partir de* "q". A pesar de resultar una versión de la paradoja del mentiroso, es quizá interesante reducir esta paradoja al Esquema de Inclusión para ver cómo una oración, o una promesa, o una

[37] A diferencia de lo que ocurre con μ = 'μ es falsa' que parece paradójica en cualquier contexto exclusivamente sobre la base de nuestras definiciones de verdad y falsedad y de la referencia de 'μ'. (Aunque, revisaremos esta afirmación en el capítulo 4, véase 4.5.3.)

regla que habitualmente no presentan ningún problema pueden aparecer a menudo en razonamientos con forma de contradicciones de inclusión debido a factores contingentes o contextuales.

Sea Ω el conjunto de todas las verdades (φ = "ser verdadero"); ψ la propiedad "ser definible"; y $\delta(X)$ (para todo X tal que $\psi(X)$ y $X \subseteq \Omega$), el enunciado "q", donde se cumplen de forma contingente las siguientes condiciones y hechos:

(I) Si x dice "p" y $p \in X \subseteq \Omega$, entonces $\neg q \in \Omega$ (i.e., $\neg q$ es verdadera).

(II) Si x dice "p" y $p \notin X \subseteq \Omega$, entonces $q \in \Omega$ (i.e., q es verdadera).

(III) x dice "q".

(En el caso que nos ocupa, "q" es el enunciado "el cocodrilo comerá al niño" y las condiciones (I) y (II) son las promesas hechas por el cocodrilo al padre: (I) "Si dices 'p' y 'p' resulta pertenecer a cierto grupo de oraciones verdaderas, X, entonces '$\neg q$' es (o será) verdad"; y (II) "Si dices 'p' y 'p' no pertenece a X, entonces 'q' es (o será) verdad". Las promesas del cocodrilo se vuelven inconsistentes en presencia de (III) sólo cuando el grupo X de oraciones es el de *todas* las verdades y (I) y (II) pasan así a ser: "Si dices 'p' y 'p' es/no es verdad, entonces...") Tenemos, pues, la siguiente contradicción de inclusión:

1) $\Omega = \{p$: 'p' es verdadera$\}$ existe y Ω es definible.

2) Si $X \subseteq \Omega$ y X es definible, $\delta(X) = q$ (i.e., $\delta(X)$ es igual a lo que dice x).

Trascendencia: Supongamos que $q \in X$, entonces $q \in \Omega$ (ya que $X \subseteq \Omega$) y, como x ha dicho "q" (por III), entonces tenemos $\neg q \in \Omega$ (por

I), lo que implica la verdad de una contradicción ya que si q y $\neg q$ pertenecen a Ω, el conjunto de todas las verdades, entonces ambas son verdaderas. Así pues, rechazamos $q \in$ X, nuestro supuesto de partida, y demostramos la condición de trascendencia: $q \notin$ X.

Clausura: Hemos visto que $q \notin$ X (por la condición de *trascendencia*) y sabemos que x ha dicho "q", por tanto (II) nos permite inferir $q \in \Omega$ demostrando así la condición de clausura.

Cuando el conjunto X de oraciones verdaderas es igual a Ω (i.e., el conjunto de todas las verdades) tenemos una contradicción apelando a (III) y a los principios (I) y (II), i.e., $q \in \Omega$ y $q \notin \Omega$. Vemos, pues, que la paradoja del cocodrilo y todas aquellas que responden a la descripción general aquí aportada encajan en el Esquema de Inclusión. En este caso, la paradoja se deshace rechazando la posibilidad de satisfacer conjuntamente los principios (I) y (II) en aquellos contextos en los que (III) es un hecho. En el caso que nos ocupa, podemos decir que la paradoja del cocodrilo se resuelve rechazando mediante una *reductio* la posibilidad de satisfacer conjuntamente las promesas (I) y (II) en *cualquier* contexto, esto es, sin introducir restricciones adecuadas.

2.10.4 La paradoja del hiper-juego [15]

[15] La paradoja del hiper-juego. En algunos juegos para dos personas donde A mueve en primer lugar y B en segundo, existe siempre una estrategia ganadora para uno de los dos contrincantes desde el principio de cualquier partida. Llamemos "cerrados" a estos juegos. El hiper-juego es un juego para dos personas con las siguientes reglas básicas: 1) El

primer movimiento de toda partida consiste en la elección, por parte de A, de un juego cerrado, *j*. 2) En el segundo movimiento B ocupa la posición de jugador A (primer jugador) en *j* y realiza un movimiento en dicho juego. 3) El hiper-juego continua como una partida normal del juego *j* hasta su final. El hiper-juego debe ser un juego cerrado, ya que cualquier juego que se elija en el primer movimiento lo es. Pero, si el hiper-juego es un juego cerrado, entonces es elegible por A en su primer movimiento, y después por B, y por A de nuevo, ... Tendríamos así una partida abierta, sin ganadores, y, consiguientemente, el hiper-juego no sería un juego cerrado. La paradoja consiste, pues, en que el hiper-juego es un juego cerrado si y sólo si no lo es.

La paradoja del hiper-juego también encaja en el Esquema de Inclusión. Para ello tomemos como φ la propiedad de "ser un juego (para dos personas) cerrado"; ψ será, como en tantos otros casos, la propiedad de "ser definible". Tenemos, pues:

1) $\Omega = \{y\text{: } y$ es un juego cerrado para 2 personas$\}$ existe y Ω es definible.
2) Si $X \subseteq \Omega$ y X es definible, entonces $\delta(X)$ = Hiper-X (H-X).

Donde H-X es un juego para dos personas definido de la siguiente manera: 1. En el primer movimiento el jugador A elige un juego *j* tal que $j \in X$. 2. En el segundo movimiento el jugador B ocupa la posición de jugador A en *j* y efectúa el primer movimiento en una partida de *j*. 3. La partida continúa como estipulan las reglas de *j* hasta su fin.

Se puede demostrar que $\delta(X)$ satisface las condiciones de "trascendencia" y "clausura" como estipula EI:

Trascendencia: Supongamos que $\delta(X) \in X$ (i.e., $H\text{-}X \in X$). Si éste es el caso, entonces el jugador A puede elegir H-X en su primer movimiento, pero entonces B puede elegir a continuación H-X y A, de nuevo, H-X, etcétera. Así pues, no siempre existe una estrategia ganadora para A o B en H-X, lo que significa que H-X no es un juego cerrado. Sin embargo, esto contradice nuestro supuesto inicial, ya que $H\text{-}X = \delta(X) \in X \subseteq \Omega$ y, por tanto, $H\text{-}X \in \Omega$ (H-X es cerrado). Concluimos que $\delta(X) \notin X$.

Clausura: Sabemos que $\delta(X) \notin X$ (i.e., $H\text{-}X \notin X$), sabemos igualmente que todos los juegos elegibles por A en H-X son cerrados, ya que A debe elegir entre los juegos de X y $X \subseteq \Omega$. Así pues, existe siempre una estrategia ganadora para A o B en cualquier partida de H-X,[38] por lo tanto, $H\text{-}X \in \Omega$, esto es, $\delta(X) \in \Omega$.

El hiper-juego es, obviamente, $H\text{-}\Omega$, es decir, $\delta(\Omega)$.

2.10.5 Funciones no enumerables y funciones no (Turing-)computables

En esta sección ofreceré dos razonamientos que habitualmente no consideramos paradójicos (como ocurría con el teorema de Cantor) pero que encajan en el Esquema de Inclusión. Se trata de dos demostraciones por *reductio*. En un caso se establece la imposibilidad de enumerar todas las funciones monádicas. En el otro, se demuestra la existencia de funciones monádicas incomputables (por máquinas de Turing).[39] Vayamos con el primer razonamiento.

Se trata de demostrar que las funciones monádicas no son enumerables (es decir, que hay *más* funciones monádicas que números naturales) y apelamos para ello a una *reductio*: Supongamos que estas funciones son enumerables. Tenemos entonces una lista exaustiva: $f_0, f_1, f_2, f_3, \ldots$ que asigna a cada función monádica un número natural exclusivo. A partir de esta lista, podemos definir una función, h, que asigne a cada número n el resultado de añadir 1 al valor numérico que la n-ava función de la lista (f_n) le asigna a n. Esto es: $h(n) = f_n(n) + 1$ (si f_n es una función parcial no definida para n, entonces $h(n) = 1$). La función h es monádica pero no puede aparecer en nuestra supuesta enumeración de *todas* las funciones monádicas. Supongamos que h aparece en la lista. Existirá, pues, un número natural n tal que $h = f_n$, de donde se sigue un absurdo: $f_n(n) = h(n) = f_n(n) + 1$. Así pues, h no puede aparecer en la lista y, por tanto, no todas las funciones monádicas son enumerables.

Bajo el supuesto –necesario para la *reductio*– de que hay tantas funciones monádicas como números naturales, podemos expresar el argumento anterior como una contradicción de inclusión. Para ello consideramos una enumeración completa de las funciones monádicas

---

[38] De hecho, existe desde el principio una estrategia ganadora para A: elegir en el primer movimiento un juego cerrado con estrategia ganadora para el segundo jugador (posición que ocupará A después en el juego).

[39] Que las funciones sean o no monádicas (i.e., funciones con un sólo argumento) es secundario. Estos mismos resultados se pueden obtener para funciones n-ádicas (i.e.,

ordenadas en forma de lista. Para simplificar las cosas, cada número natural, n, será tratado como el nombre de la n-ava función de la lista. A continuación interpretamos las variables de EI como sigue: φ es la propiedad "ser el código (i.e., el número natural) que nombra a una función monádica" –bajo el supuesto problemático Ω es, a la vez, el conjunto de todos los números naturales y el de *todas* las funciones monádicas–; ψ es una propiedad universal: "ser idéntico a sí mismo"; y δ es la función que a cada $X \subseteq \Omega$ le asigna el *código*, m, de una función monádica h definida del siguiente modo:

(h1) $h(n) = 1$, si $n \notin X$;

(h2) $h(n) = 1$, si $n \in X$ pero la función cuyo código es n (i.e., f_n) no asigna un valor a n;

(h3) $h(n) = f_n(n) + 1$, si $n \in X$ y f_n asigna un valor a n.

Trascendencia: El código, $\delta(X) = m$, de la función h no puede pertenecer a X. Si $m \in X$, entonces $h = f_m$ y tenemos un absurdo: $f_m(m) = h(m) = f_m(m) + 1$ (como h siempre está definida no hace falta considerar el caso h2). Así pues, $m \notin X$ y, por tanto, $\delta(X) \notin X$.

Clausura: Como h es una función monádica y hemos supuesto (para una *reductio*) que todas las funciones monádicas se pueden codificar mediante números naturales, el código de h pertenece a Ω, i.e., $\delta(X) \in \Omega$.

Otra *reductio* que podemos presentar como una contradicción de inclusión es la que muestra que existen funciones monádicas no

---

con n argumentos, donde n es un número natural arbitrario). Nos centraremos en funciones monádicas por razones de simplicidad expositiva.

computables por máquinas de Turing.[40] Dado que "ser computable por una máquina de Turing" y "ser computable por recursión" son conceptos equivalentes (*C&L*, caps. 6-8), esta *reductio* muestra que hay funciones no computables por recursión.[41]

El argumento que establece este resultado es prácticamente idéntico al que acabamos de dar al hablar de la no enumerabilidad de las funciones monádicas. Se puede demostrar (véase *C&L*, cap. 5) que las funciones computables por máquinas de Turing (i.e., las funciones recursivas) son *enumerables*. Ahora bien, si esto es así, podemos construir una función monádica, ***h'***, del mismo modo en que construimos la función ***h*** del ejemplo anterior. Esta vez, definimos ***h'*** a partir de la enumeración de las funciones (Turing-)computables y, por el mismo razonamiento, concluimos que si *h'* es (Turing-)computable (i.e., si aparece en la enumeración), entonces existe una función (Turing-)computable f_n tal que $h' = f_n$, de donde se sigue (por la definición de *h'*) un absurdo: $f_n(n) = h'(n) = f_n(n) + 1$. Este razonamiento encaja en el Esquema de Inclusión del mismo modo que el anterior. La única diferencia es que ahora, en nuestra interpretación de Ω, φ aparece como la propiedad: "ser el código (i.e., el número natural) que nombra a una función *computable por una máquina de Turing*" y Ω se puede entender

[40] Todos los resultados referidos en esta sección se pueden encontrar en cualquier manual de teoría de la computabilidad. Usaremos aquí Boolos y Jeffrey 1989, que abreviaré como *C&L* seguido del capítulo (cap.) o páginas de referencia. El teorema que abordaremos ahora, por ejemplo, aparece en *C&L*, cap. 5.

[41] Según la tesis de Church (no demostrada), una función es computable si, y sólo si, es computable por una máquina de Turing. Si la tesis es correcta, la función que definiremos a continuación es incomputable en cualquier sentido. Si no lo es, podría ser computable, pero no por una máquina de Turing o por recursión.

a la vez como el conjunto de las funciones computables por máquinas de Turing o como el conjunto de los números naturales. Todo lo demás (la interpretación de δ y ψ) permanece igual teniendo en cuenta el nuevo significado de φ. Las demostraciones de trascendencia y clausura son las mismas de antes, aunque la premisa que rechaza la *reductio* en este caso es que $\delta(\Omega)$ sea una función Turing-computable. Esto es, se rechaza $\delta(\Omega) \in \Omega$ y no la existencia de Ω, i.e., de una enumeración de las funciones (Turing-)computables.

2.10.6 La paradoja de Curry [17]

[17] La paradoja de Curry (o de Löb). Existen muchas variantes de esta paradoja, las más simples se construyen a partir de oraciones del tipo:

(*s*) Si *s* es verdadera, entonces p.

donde 'p' puede ser cualquier oración: 'Dios existe', 'Dios no existe', '2 = 3', ... La oración *s* dice de sí misma que si es verdadera, entonces es el caso que p. El problema que plantea *s* es que a partir de ella es posible demostrar p, es decir, cualquier cosa –incluyendo contradicciones o falsedades. Veamos como.

1. *s* es verdadera	(suposición)
2. *s* es verdadera $\rightarrow$ p	(Esquema-T: "x es verdadera $\leftrightarrow$ p" + def. *s*, en 1)
3. p	(Modus Ponens, 1, 2)
4. *s* es verdadera $\rightarrow$ p	(Introducción $\rightarrow$, 1-3)
5. *s* es verdadera	(Esquema-T: + def. *s*, en 4)
6. p	(Modus Ponens, 4, 5)

Como 'p' puede ser cualquier cosa, podemos demostrar un absurdo. Podemos representar s del siguiente modo: s = '$s \in T \rightarrow \perp$', donde T es la colección de todas las verdades y $\perp$ una falsedad lógica. La paradoja de Curry muestra que a partir de s podemos demostrar $\perp$.

¿Encaja la paradoja de Curry en EI? Si interpretamos '$\rightarrow$' en s como un condicional material (como, de hecho, ocurre en lógica clásica) y no prescindimos de la negación, la respuesta es sí. La paradoja de Curry es una contradicción de inclusión del tipo [$Y \subset \wp(\Omega)$ y $\Omega \in Y$]. En este caso tenemos φ = "ser verdadera" y ψ = "ser definible".

1) $\Omega = T$ = {y: y es verdadera} existe y T es definible.

2) Si $X \subseteq \Omega$ y X es definible, $\delta(X) = s$ = '$s \in X \rightarrow \perp$'

Trascendencia: $\delta(X) \notin X$

1. $\delta(X) \in X$	(Suposición)
2. $s \in X$	(definición δ: 1)
3. $s \in T$	($X \subseteq T$: 2)
4. $s \in X \rightarrow \perp$	(def. s + Esquema-T: 3)
5. $\perp$	(Mod. Pon.: 2, 4)
6. $\delta(X) \notin X$	(Absurdo: 1-5)

Clausura: $\delta(X) \in \Omega$

1. $s \in X$	(suposición)
2. $s \in T$	($X \subseteq T$: 1)
3. $s \in X \rightarrow \perp$	(def. s + Esquema-T: 2)
4. $\perp$	(Modus Ponens: 1, 3)
5. $s \in X \rightarrow \perp$	(Introduc $\rightarrow$: 1-4)
6. $s \in T$	(Esquema-T: 5)
7. $\delta(X) \in T = \Omega$	(def. s: 6)

Una vez más, la contradicción se produce cuando X = T, ya que $\delta(T) = s$ = '$s \in T \rightarrow \perp$', $\delta(T)$ es la paradoja de Curry. Así pues, la paradoja de Curry también encaja en EI si leemos '$\rightarrow$' como un condicional material. Véase, no obstante, la discusión en 2.8.1.

3 Cómo interpretar EI: diagonalizadores y totalidades absolutas

3.1 Introducción

Hemos visto en el capítulo anterior que cierto número de razonamientos cuya conclusión es una contradicción pueden ser descritos de forma genérica mediante el Esquema de Inclusión (2.5). Este esquema tiene su origen en unas observaciones de Russell acerca de la identidad estructural de las paradojas "conjuntistas" más famosas y ha sido posteriormente modificado por Priest para poder encajar en él lo que Ramsey dio en llamar "paradojas semánticas". Siguiendo a Priest, hemos denominado "contradicciones de inclusión" a este tipo de razonamientos examinando algunos de ellos. Aparte de las paradojas más relevantes que Priest reduce a EI, hemos descrito en el anterior capítulo tres paradojas (no consideradas por él) que también encajan en EI: la paradoja de la cita (un caso límite de EI), la paradoja de la denotación de Barwise y Moss, la del hiper-juego. Igualmente, hemos comprobado mediante dos ejemplos (el teorema de Cantor y el de la no enumerabilidad de las funciones monádicas) que ciertos fragmentos de teoremas bien conocidos pueden considerarse contradicciones de inclusión. Este último hecho refuerza la idea de que el esquema de Priest proporciona fundamentalmente una descripción estructural de un tipo determinado de *razonamientos* con independencia de que éstos se perciban o no como paradójicos. Por otra parte, EI no pretende ofrecer una descripción general de *todas* las paradojas en general sino sólo de aquellas en las que los elementos que forman parte del citado esquema juegan un papel crucial en la demostración de una contradicción. En este sentido hemos argumentado con Priest que el único supuesto contraejemplo a su esquema, i.e., la paradoja de Curry sin negación, no puede ser una contradicción de inclusión, al no poder haber *contradicción* sin *negación*. Aunque esta versión de la paradoja ha de pertenecer forzosamente a una familia distinta de paradojas, hemos visto que en cuanto podemos introducir una negación en una prueba por *reductio* ($\alpha \rightarrow \perp \Rightarrow \neg\alpha$) para expresar la paradoja mediante una contradicción, la paradoja se convierte de inmediato en una *contradicción* de inclusión.

Hasta ahora nos hemos limitado a comprobar que una serie de razonamientos encajan formalmente en un esquema, EI, pero cabe preguntarse a continuación cuál es el valor real de este hecho: ¿Nos ayuda a comprender mejor las paradojas y su relación con otras contradicciones de inclusión? Ésta es la cuestión que abordaremos en el presente capítulo. Nuestra preocupación aquí será doble: 1) Por un lado, buscamos una interpretación de los elementos de EI (δ, Ω, φ y ψ) que arroje luz sobre las paradojas y nos ayude a comprender cómo y por qué se producen; 2) por otro, deseamos mostrar en qué sentido muchas de las intuiciones y descripciones que se asocian tradicionalmente a las paradojas reflexivas tienen perfecta cabida en la interpretación de EI aquí propuesta.

3.2 ¿Qué significado tiene el Esquema de Inclusión? Los límites del pensamiento

En *Beyond the Limits of Thought* (en lo sucesivo, *BLT*), Graham Priest propone mucho más que un simple esquema general en el que encajar todas las contradicciones de inclusión. A través de un amplio recorrido histórico, Priest se detiene en una serie de proyectos recurrentes que han marcado el ámbito de reflexión de la filosofía. Todos ellos tienen algo en común, son proyectos que pretenden fundamentar y explorar los límites de aplicación de ciertos conceptos básicos pero, en cierto modo, son también proyectos que empujan esos conceptos más allá de sus límites. La mayoría de ideas que expondremos a continuación son desarrollos de ideas que podemos encontrar de forma más o menos explícita en *BLT*, aunque mi presentación no coincide plenamente con la de Priest.

3.2.1 Generadores iterativos, ω-secuencias y secuencias transfinitas

Gran parte de los conceptos que utilizamos para explicar y describir el mundo sugieren la formación de secuencias mediante operaciones iterativas. Cuando nos preocupamos por establecer las causas de un suceso, S_0, por ejemplo, generamos una cadena causal a partir de S_0 iterando el operador "ser causado por": <S_0 es causado por S_1 es causado por S_2 es ...>; igualmente, cuando queremos describir la composición de un complejo, *C*, generamos un análisis que va encadenando sus elementos mediante el operador "estar constituido por". Priest denomina *generador* a cualquier tipo de operador cuya aplicación reiterada produce una de estas secuencias. Cada vez que aplicamos un generador, *g*, sobre un objeto, *x*, de cierto tipo pueden ocurrir dos cosas: o bien no obtenemos nada, o bien obtenemos un objeto, $g(x)$, del mismo tipo que *x* sobre el que podemos volver a aplicar *g*. El resultado de iterar *g* es una secuencia del tipo $\langle x, g(x), g(g(x)), ...\rangle$ en la que cada miembro (salvo el primero) es definido a partir del anterior mediante la ecuación: $x_{n+1} = g(x_n)$ $(0 \leq n < \omega)$.[1] Llamaremos *ω-generador* a todo operador cuya iteración origine una secuencia de esta clase.[2] La definición conjuntista de la función recursiva "sucesor de" (*suc*), según la cual el sucesor de un conjunto x_n es el resultado de añadir los miembros de x_n al conjunto $\{x_n\}$ es un ejemplo paradigmático de *ω-generador*: $x_{n+1} = suc(x_n) = x_n \cup \{x_n\}$. Partiendo de $\varnothing$, *suc* genera la secuencia:

$$\langle x_0 = \varnothing, x_1 = \{\varnothing\}, x_2 = \{\varnothing, \{\varnothing\}\}, x_3 = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}, ...\rangle.$$ [3]

Como es bien sabido, no todas las secuencias pueden formarse de este modo. Consideremos el siguiente caso: $r = \langle x_0, x_1, x_2, ..., x_\omega\rangle$. Ni *suc* ni cualquier otro operador iterativo, *g*, cuyos valores se definan mediante la ecuación $x_{n+1} = g(x_n)$ podría generar *r*. Al venir precedido por una serie infinita de elementos, x_ω no puede definirse a partir de su predecesor

---

[1] "Suppose that we have an operator that applies to objects of a certain kind. When applied to an object of that kind it may produce nothing, or it may produce an object of the same kind. I will call such an operator a *generator* since, by applying the operator again and again, we can generate a sequence of objects. Whatever results by iterating applications of the operator as often as possible in this way I will call a *limit of iteration.*" (*BLT*, p. 26)

[2] La secuencia no tiene por qué ser infinita, si $g(x_n)$ no produjese nada, entonces *g* formaría una secuencia finita $s = \langle x_0, x_1, ..., x_n\rangle$.

[3] La definición no conjuntista de la operación aritmética "sucesor de", ***suc***, es otro ejemplo de ω-generador: ***suc***(n+1) = s(n). A partir de 0, ***suc*** genera la secuencia de los números naturales: <0, s(0) = 1, s(s(0)) = 2, s(s(s(o))) = 3, ...>.

inmediato, tal predecesor *no existe*, no existe ningún, x_n, en r tal que $x_\omega = g(x_n)$. Las secuencias como r en las que aparece algo a continuación de una sucesión infinita de elementos reciben el nombre de secuencias *transfinitas* y no pueden ser formadas por *ω-generadores*. En estas secuencias lo más apropiado es definir cada elemento a partir de *todos* sus predecesores (que podrían ser infinitos en número y no tener un último elemento). Llamaremos *generador transfinito* a cualquier operador iterativo, g, que origine una secuencia ordenada, $s = \langle x_0, x_1, x_2, ..., x_\omega, x_{\omega+1}, ...\rangle$, en la que cada uno de sus miembros, x_α (incluido x_0), se defina a partir del conjunto de *todos* sus predecesores mediante la ecuación: $x_\alpha = g(\{x_\beta: \beta < \alpha\})$, donde α y β son números ordinales. Cabe observar que, al igual que los *ω-generadores*, los generadores transfinitos también actúan siempre sobre (conjuntos de) cosas que han producido previamente.[4] Un ejemplo paradigmático de generador transfinito es $\bigcup_{\beta<\alpha} suc(x_\beta)$. Este operador complejo toma todos los conjuntos x_β (con $\beta < \alpha$) y aplica a cada uno de ellos el operador *suc*, a continuación, efectúa la unión de todos los conjuntos sucesores así obtenidos. El resultado de iterar este operador es una secuencia transfinita en la que cada elemento se define mediante la ecuación $x_\alpha = \bigcup\{suc(x_\beta): \beta < \alpha\}$:

$$x_0 = \bigcup\{suc(x_\beta): \beta < 0\} = \bigcup\{\varnothing\} = \varnothing$$

$$x_1 = \bigcup\{suc(x_0)\} = \bigcup\{x_0 \cup \{x_0\}\} = \bigcup\{\varnothing \cup \{\varnothing\}\} = \bigcup\{\{\varnothing\}\} = \{\varnothing\}$$

$$x_2 = \bigcup\{suc(x_0), suc(x_1)\} = suc(x_0) \cup suc(x_1) = \{\varnothing, \{\varnothing\}\}, ...$$

el primer elemento transfinito es:

$$x_\omega = \bigcup\{suc(x_\beta): \beta < \omega\} = x_1 \cup x_2 \cup x_3 \cup ... = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, ...\}, ...$$

---

[4] Priest amplia el término "generador" de este modo para dar cuenta de series

Llamaremos '*tsuc*' al operador $\bigcup_{\beta<\alpha} suc(x_\beta)$. *tsuc* es, de hecho, una generalización transfinita de *suc*, i.e., un operador que continúa o *trasciende* las series infinitas formadas por *suc*. Para todo elemento de las series generadas por *tsuc* y *suc* entre x_0 y x_ω (excluyendo éste último), se cumple que $x_{n+1} = suc(x_n) = tsuc(\{x_m: m < n+1\})$.[5]

3.2.2 Límites iterativos y totalidades absolutas

Cuando pensamos en un generador, *g*, y en la serie o cadena que forma, pensamos también en sus "límites", es decir, en lo que puede pertenecer a la serie y en lo que debe quedar fuera de ella o, desde otra perspectiva, en las series que puede formar *g* y en las que no. La propia naturaleza de un generador determina en gran medida cuáles son sus "límites". A menos que el operador considerado fuese capaz de ordenar secuencialmente "todo" lo que hay en el "universo" (si es que esa idea tiene sentido), concebimos habitualmente el límite de un generador, o de una secuencia, como algo que ha de estar restringido de algún modo. Por ejemplo, el operador ***suc*** de la nota 3 no puede generar ni series con

---

transfinitas. Véase la nota 15.

[5] Otro ejemplo de generador transfinito es $\bigcup_{\beta<\alpha} \wp(x_\beta)$, al que llamaremos para abreviar '*u-sub*'. *u-sub*$(\{x_\beta: \beta < \alpha\})$ une los conjuntos potencia de todo x_β (con $\beta < \alpha$). La secuencia generada responde a la ecuación $x_\alpha = \bigcup\{\wp(x_\beta): \beta < \alpha\}$:

$$x_0 = \bigcup\{\wp(x_\beta): \beta < 0\} = \bigcup\{\varnothing\} = \varnothing,$$
$$x_1 = \bigcup\{\wp(x_0)\} = \bigcup\{\wp(\varnothing)\} = \bigcup\{\{\varnothing\}\} = \{\varnothing\},$$
$$x_2 = \{\varnothing, \{\varnothing\}\},$$
$$x_3 = \{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\varnothing, \{\varnothing\}\}\},$$
$$\ldots$$
$$x_\omega = \bigcup\{\wp(x_\beta): \beta < \omega\} = x_1 \cup x_2 \cup x_3 \cup \ldots = \{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\varnothing, \{\varnothing\}\} \ldots\},$$
$$\ldots$$

Un generador parecido es ***u-sub***$(\{y: y \in x\}) = \bigcup\{\wp(y): y \in x\}$. Cuando ***u-sub*** parte de $\varnothing$, las series generadas por *u-sub* y ***u-sub*** son la misma.

elementos transfinitos ni series en las que aparezcan "cabras", sólo puede generar series de números naturales. De igual modo, una cadena causal (o analítica) formada a partir de un eslabón inicial, *x*, mediante la iteración de algún operador que asigna a cada eslabón sus causas (o sus elementos) *sólo* puede contener cosas que sean causas (o elementos) relacionados con *x*. Su límite es, pues, el conjunto de todo aquello que interviene en la causación (o composición) de *x*. En cierto sentido, pensar en el "límite" de una serie es, pues, pensar en un conjunto de elementos que, por su naturaleza, por satisfacer cierta propiedad o propiedades P, pueden ser generados por un operador de cierto tipo.

Pero, pensar en el "límite" de una serie u operador tiene también el matiz de pensar en un "límite iterativo", en la posibilidad de alcanzar un estadio en la iteración de un operador en el que nada nuevo se añada a la serie ya generada. Pensar, por ejemplo, en el límite de las cadenas causales o analíticas mencionadas, es pensar en la posibilidad de un estadio finito, infinito o transfinito en el que una nueva iteración del operador que las genera no añada nada más –una causa ulterior o elemento– a la cadena ya existente. Uniendo ambas ideas, tenemos que el límite de una serie, o cadena, formada mediante un operador *g* se puede identificar de modo natural con el conjunto de todos los *x* de cierto tipo (todos "los Ps") que han sido generados por *g* antes de alcanzar un estadio en el que cualquier ulterior aplicación de *g* es estéril. El conjunto así delimitado puede tener un número finito o infinito de elementos y puede tener un último elemento (finito o transfinito) o carecer de él.

El límite ideal de una operación iterable *f* se nos presenta, pues, asociado a una *totalidad absoluta*, X, de cosas que satisfacen cierta(s) propiedad(es) P. Decir que X es una totalidad absoluta en este contexto, no significa decir que no haya totalidades mayores o totalidades que contengan a X (ni siquiera que X sea una colección infinita), significa, simplemente, decir que no puede haber totalidades mayores que X formadas *exclusivamente* a partir de *f*. X es una totalidad absoluta porque determina el *límite absoluto* en la iteración de *f*, cualquier aplicación de *f* produce un elemento de X. La totalidad, DOn, de los números ordinales definibles, por ejemplo, está incluida en la totalidad mayor de los ordinales (On), sin embargo DOn es el resultado absoluto de aplicar operaciones de definición sobre elementos de On.[6] Ninguna operación cuyo resultado fuese la definición de un número ordinal podría agrandar DOn, de lo contrario DOn no contendría a *todos* los números ordinales definibles. Del mismo modo, un átomo forma parte de innumerables compuestos que lo contienen. Sin embargo, la idea de "átomo" (físico, conceptual, etc.) está ligada a la totalidad absoluta de divisiones (físicas, etc.) a las que uno puede someter un objeto (físico, etc.). El límite iterativo de un generador, *g*, viene delimitado, pues, por la totalidad absoluta de cosas que puede producir iterativamanete *g* antes de alcanzar cierto estadio más allá del cual ninguna iteración de *g* aumenta la serie ya formada.

[6] No tiene por qué actuar una sola operación iterable, *f*, en la construcción de una de

3.2.3 Iteración y orden

Hemos hablado de generadores, de las series que originan y de sus límites. Ahora bien, hablar de series es hablar de *orden*, de conjuntos cuyos elementos pueden ser *ordenados* por medio de una relación de cierto tipo. Pero, ¿qué es una relación de orden, cómo podemos determinar qué es una serie o qué cosas se pueden ordenar en series? Aunque estas cuestiones puedan resultar muy básicas será de utilidad detenerse un instante en ellas con vistas a posteriores discusiones. La definición de orden que resultará relevante en el presente trabajo es la definición rigurosa y ya familiar que manejamos en teoría de conjuntos. En este contexto, es común afirmar que un conjunto X está *parcialmente ordenado* cuando sus miembros mantienen una relación diádica, R, habitualmente representada mediante '≤', que es *reflexiva*, *antisimétrica* y *transitiva*. Esto es: $\forall x, y, z \in$ X, (1) $x \leq x$; (2) $(x \leq y \wedge y \leq x) \rightarrow x = y$; (3) $(x \leq y \wedge y \leq z) \rightarrow x \leq z$. Un conjunto X parcialmente ordenado por ≤ se representa mediante (X, ≤).

Cuando se cumple para todo $x, y \in$ X que, o $x \leq y$, o $y \leq x$, se dice que X está *totalmente ordenado* por ≤, y si ocurre, además, que todo subconjunto no vacío, Y, de X tiene un *menor* (o *primer*) elemento, i.e., un $x \in$ Y tal que, para todo $y \in$ Y, $x \leq y$, entonces decimos de X y de ≤ que están *bien fundamentados*. (Trivialmente, todo conjunto bien fundamentado está *totalmente ordenado*, ya que todos sus subconjuntos $\{x, y\}$ tienen un menor elemento.) Un conjunto y una relación bien

---

estas totalidades.

fundamentados reciben el nombre de conjunto *bien ordenado* y de relación de *buen orden*, respectivamente.

El vínculo entre órdenes y series es muy fuerte. Decir que un conjunto puede ser bien ordenado es equivalente a decir que podemos formar una serie con sus elementos mediante algún tipo de generador. Para ver esto con claridad, imaginemos un conjunto, X, bien ordenado y no vacío. Sabemos que X tiene un *primer* elemento, x_0, con el que podemos formar una serie singular, $\{x_0\}$. Si le restamos $\{x_0\}$ a X, obtendremos X – $\{x_0\}$, un subconjunto de X que, si no es vacío, tendrá también un primer elemento por estar X bien ordenado. Este elemento será menor que todos los elementos de X distintos de x_0, así pues, podemos considerarlo "el *segundo* elemento de X", x_1, y formar otra serie $\{x_0, x_1\}$. Si restamos de X esta nueva serie, X – $\{x_0, x_1\} \subseteq$ X, y el subconjunto resultante no es vacío, tendrá un menor elemento (el tercero de X), x_2, y podemos volver a repetir la operación una y otra vez. Este proceso genera una serie, $\{x_0, x_1, x_2, ...\}$, que puede tener elementos transfinitos, ya que, si X – $\{x_0, x_1, x_2, ...\} \subseteq$ X no es vacío, tendrá un menor elemento, x_ω. Cuando alcancemos un estadio en el que X – $\{x_0, x_1, ..., x_\omega, ...\}$ sea vacío, la serie restada a X tendrá los *mismos* elementos que X y, por tanto, será el resultado de convertir el conjunto bien ordenado X en una serie. El generador, *g*, que produce los elementos de esta serie actúa sobre subconjuntos de X y se puede definir del siguiente modo: Para todo Y $\subseteq$ X, *g*(Y) = "el menor elemento de X (según $\leq$) que no pertenece a Y" (i.e., "el menor elemento de X – Y"). Como X está bien ordenado, el subconjunto X – Y tiene (siempre que Y $\neq$ X) un menor elemento $g(Y)$. Cuando Y = X, g alcanza su límite iterativo. Los elementos, x, de la serie formada por g satisfacen la ecuación $x = g(\{y \in X: y \leq x \wedge y \neq x\})$.[7]

Desde el extremo opuesto tenemos que, partiendo de una serie $X = \{x_0, x_1, ..., x_\omega, ...\}$ construida mediante un generador, g, obtenemos de forma inmediata una relación de buen orden en X. Dado que g satisface la ecuación $x_\alpha = g(\{x_\beta: \beta < \alpha\})$ (para todo $x_\alpha \in X$, donde α y β son ordinales), podemos construir un buen orden sobre X mediante la siguiente estipulación: $g(\{x_\beta: \beta < \alpha\}) \leq y$, para todo $y \in X - \{x_\beta: \beta < \alpha\}$:

$g(\varnothing) = x_0$ (0) $\forall y \in X - \varnothing, x_0 \leq y$;

$g(\{x_0\}) = x_1$ (1) $\forall y \in X - \{x_0\}, x_1 \leq y$;

$g(\{x_0, x_1\}) = x_2$ (2) $\forall y \in X - \{x_0, x_1\}, x_2 \leq y$;

...

$g(\{x_0, x_1, x_2, ...\}) = x_\omega$ (ω) $\forall y \in X - \{x_0, x_1, x_2, ...\}, x_\omega \leq y$;

...

Obviamente, $\leq$ es un buen orden sobre X ya que es una relación reflexiva, antisimétrica y transitiva y si $Y \subseteq X$, entonces Y tiene un menor elemento x_α, donde α es el menor ordinal que aparece como subíndice de un elemento de Y.

Así pues, podemos considerar equivalentes las nociones de "serie formada por un generador" y "conjunto bien ordenado". Sin embargo, aún queda algo por aclarar. En ambas transiciones (orden-serie y serie-

[7] Cuando el operador actúa sobre $\varnothing$ tenemos "el menor elemento de X que no está en $\varnothing$", esto es, el primer elemento de X: $g(\varnothing) = x_0$.

orden) hemos utilizado una noción que todavía no hemos explicado: la noción de ordinal.

3.2.4 La serie de los ordinales, On

¿Cómo podemos precisar la idea de "serie formada por un operador iterativo"? ¿Qué series podemos concebir como resultado de la iteración de un operador, cuáles son los límites concebibles de la iteración y, por tanto, del buen orden? Una posible manera de responder a estas preguntas es construir un *modelo* de serie, *s*, y describir o definir cualquier otra serie por referencia a ella. Un conjunto X será una serie en la medida en que pueda establecerse una correlación uno a uno entre X y algún segmento inicial *s*' de la serie modelo *s* (*s*' está incluida en *s*). La relación debe ser biunívoca, i.e., a cada elemento de X le corresponderá uno de *s*' y viceversa. De este modo podemos ordenar X indicando cada uno de sus elementos mediante el elemento correspondiente de *s*'. Sólo consideraremos como series a conjuntos que cumplan este requisito. Por otra parte, la serie modelo elegida deberá ayudarnos no sólo a formar series finitas e infinitas, sino también series que contengan elementos transfinitos. La serie que buscamos existe de hecho, pues es la serie de los números *ordinales* y recibe el nombre de On.

Una de las cosas que hemos de asegurar al definir los ordinales, dado el estrecho vínculo existente entre las nociones de orden y serie, es que los elementos de On formen un buen orden según cierta relación. La definición de ordinal que actualmente utilizamos se debe a von Neumann y sugiere (a grandes rasgos) que un ordinal α sea identificado con el conjunto de *todos los ordinales que preceden a* α, siguiendo cierta relación, $\leq$, de buen orden. Esta definición se puede hacer más precisa. Consideremos un conjunto bien ordenado $(X, \leq)$ y tomemos uno de sus elementos $a \in X$. A partir de a podemos formar un conjunto bien ordenado $X_a \subseteq X$ con todos los elementos de X que preceden a a según $\leq$: $X_a = \{x \in X: x < a\}$. Este conjunto recibe el nombre de *segmento inicial de* X (dominado por a). Podemos definir ahora un ordinal como un conjunto bien ordenado $(X, \leq)$ tal que, si $a \in X$, entonces $a = X_a$. (a es el conjunto de todos sus predecesores en X según $\leq$). La definición de von Neumann es un hallazgo afortunado por muchos motivos. A partir de ella podemos inferir que todo ordinal es igual al conjunto de los ordinales que lo preceden y podemos demostrar, además, las siguientes equivalencias: Si α y β son ordinales, entonces $\alpha < \beta \leftrightarrow \alpha \in \beta$ ($\leftrightarrow \alpha \subset \beta$) y $\alpha \leq \beta \leftrightarrow \alpha \subseteq \beta$. Estas afirmaciones implican que, en el caso de los ordinales, es posible identificar el orden '$\leq$' con '$\subseteq$' y el orden estricto '$<$' con '$\in$' (o con '$\subset$').[8] La relación $\subseteq$ constituye además un *buen orden* en On ya que es reflexiva, antisimétrica y transitiva y podemos demostrar, dado cualquier conjunto de ordinales X, que su intersección, $\cap X$, es el menor elemento de X según $\subseteq$, esto es:

---

[8] Dados tres ordinales cualesquiera α, β y γ, siempre se cumple que:
1) $\alpha \subseteq \alpha$
2) $(\alpha \subseteq \beta \wedge \beta \subseteq \alpha) \rightarrow \alpha = \beta$
3) $(\alpha \subseteq \beta \wedge \beta \subseteq \gamma) \rightarrow \alpha \subseteq \gamma$
Si sustituimos '$\subseteq$' por '$\in$', tenemos un orden *estricto* en On. La relación '$\in$' es *no-reflexiva*, *asimétrica* y *transitiva*:
1) $\alpha \notin \alpha$
2) $\alpha \in \beta \rightarrow \beta \notin \alpha$
3) $(\alpha \in \beta \wedge \beta \in \gamma) \rightarrow \alpha \in \gamma$.

1) $\cap X$ es un ordinal

2) $\cap X \in X$

3) $\cap X \subseteq \alpha$, para todo $\alpha \in X$

Una característica muy relevante de los ordinales es que son conjuntos *$\in$-transitivos*: dados tres ordinales α, β y γ, se puede demostrar que $(\alpha \in \beta \wedge \beta \in \gamma) \rightarrow \alpha \in \gamma$; o, lo que es lo mismo, $\alpha \in \beta \rightarrow \alpha \subseteq \beta$, pero no a la inversa.[9] Es más, cualquier ordinal, es a la vez un *subconjunto* de On: $\alpha \in \mathrm{On} \rightarrow \alpha \subseteq \mathrm{On}$ (pero no a la inversa), lo que implica que todo ordinal α es un conjunto de ordinales y, a la vez, un modelo de serie ordenada (hasta α): $\alpha = X_\alpha \subseteq \mathrm{On}$. Otras características relevantes de los ordinales son que la unión o intersección de un conjunto de ordinales es siempre un ordinal. La intersección de cualquier ordinal α, además, es igual a cero ($\cap\alpha = \varnothing$) y su unión igual a α ($\mathrm{U}\alpha = \alpha$).

Los ordinales pueden clasificarse en dos grupos: ordinales *sucesores* y ordinales *límite*. Un ordinal sucesor puede generarse a partir de su predecesor inmediato. Si α es un ordinal sucesor, entonces existe un ordinal β que lo precede, tal que α es el resultado de reunir en un conjunto a β y los miembros de β: $\alpha = \beta \cup \{\beta\}$. Si α no tiene predecesor inmediato, entonces representa el límite transfinito de una sucesión infinita de ordinales que lo preceden y se genera mediante la unión de todos ellos: $\alpha = \mathrm{U}_{\beta<\alpha}\, \beta$. De todos los ordinales, el primero y menor es 0, que se identifica con $\varnothing$ (el conjunto vacío). El resto de ordinales se puede obtener a partir de $\varnothing$ mediante las operaciones descritas o simplemente iterando el generador transfinito $\bigcup_{\beta\subset\alpha}$ *suc*(β) (i.e., "la unión de los sucesores de conjuntos contenidos en -o pertenecientes a- α": *u-suc*(α), para abreviar). Aquí, α se puede entender a la vez como un ordinal o como "el conjunto de todo lo generado previamente mediante *u-suc*".[10] Cada elemento (cada ordinal) de la secuencia se determina mediante la ecuación: $\alpha = \bigcup\{\beta \cup \{\beta\}: \beta \subset \alpha\}$.[11] Si el primer conjunto de ordinales sobre el que aplicamos *u-suc*(α) es ∅, la iteración transfinita del citado operador da origen a la secuencia de los ordinales, On = $\langle 0, 1, 2, ..., \omega, \omega+1, \omega+2, ..., \omega 2, \omega 2+1, ..., \omega^2, ...\rangle$. Cada ordinal transfinito, λ, se puede interpretar, en cierto sentido, como el límite de iteración de un ω-generador que parte de 0 o del ordinal transfinito inmediatamente anterior a λ, *u-suc* nos permite transcender ese límite y continuar la serie con un nuevo elemento. Desde este punto de vista, On se puede interpretar como el *límite absoluto* de la iteración de *u-suc*. Como On es la serie modelo que buscábamos, no puede haber secuencias más largas que la que determina *u-suc* partiendo de 0, nada de lo que podamos ordenar puede exceder los límites de la clase de *todos* los ordinales.

---

[9] Si $\gamma \in \alpha$, entonces γ es un ordinal menor que α y, por tanto, menor que β, así que $\gamma \in \beta$.

[10] Los ordinales finitos se corresponden con los números naturales y son ordinales sucesores: $0 = \varnothing$, $1 = \{0\}$, $2 = \{0, 1\}$, $3 = \{0, 1, 2\}$, etc.; el primer ordinal transfinito es $\omega = \{\alpha: \alpha \text{ es un número natural}\} = \bigcup_{\alpha<\omega} \alpha$.

[11] En el estadio inicial α es ∅, por tanto, $\bigcup\{suc(\beta): \beta \subset \varnothing\} = \varnothing = 0$; en el siguiente estadio α es {0} y $\bigcup\{suc(\beta): \beta \subset \{0\}\} = \{\varnothing\} = 1$; en el tercero, α es {0,1} y $\bigcup\{suc(\beta): \beta \subset \{0, 1\}\} = \{\varnothing, \{\varnothing\}\} = 2$. Como se puede apreciar, *u-suc*(α) es, básicamente, el generador *tsuc* antes descrito. La única diferencia radica en que, al utilizar ordinales como subíndices, la definición de *tsuc* presuponía el significado de "ordinal" que sólo ahora hemos especificado. Con todo, si identificamos x_0 con ∅ (como hicimos en los ejemplos de 3.2.1), la serie generada por *tsuc* es la serie de todos los ordinales.

Mediante el generador *u-suc* podemos definir, además, un nuevo generador de On que nos resultará de más utilidad en posteriores secciones. Partiendo de *u-suc*, y siguiendo las pautas descritas en 3.2.3, podemos definir primero un buen orden sobre On y, sobre la base de ese buen orden, un generador tal que, dado un conjunto de ordinales cualquiera, X, calcule siempre *el menor elemento de* On – X. Para abreviar, representaremos este generador como m.o.n(X) ("menor ordinal no en X"). Si X = {0, 1, 2, 9}, entonces m.o.n({0, 1, 2, 9}) = 3.[12]

Las ventajas de tener un modelo de orden como On se ponen de manifiesto al definir conjuntos o al comprobar si los elementos de un conjunto X ya definido satisfacen o no cierta propiedad. Si X es un conjunto pequeño, ello no supondrá un problema. Sin embargo, si X es considerablemente grande o infinito, la cosa cambia. Ordenar los elementos de X en una serie significa equiparar X con algún ordinal α asignando un ordinal $\beta \in \alpha$ a cada elemento de X. De ese modo podemos definir uno a uno todos los elementos de X o, si X ya está definido, recorrerlos ordenadamente para ver si, efectivamente, todos satisfacen cierta propiedad P. El principio de definición por recursión ordinal y los principios de verificación inductiva, se basan en la posibilidad de ordenar bien un conjunto (i.e., de formar una serie con sus elementos). En el caso de la inducción, si los elementos de X forman una ω-serie o una serie

[12] La ventaja que presenta m.o.n(X) sobre $\bigcup_{\beta \subset \alpha}$ *suc*(β) es que, dado un conjunto cualquiera X de ordinales, por ejemplo X = {1, 3, 5}, la iteración de m.o.n(X) nos permite reconstruir una serie con *todos* los elementos de On que no aparecen en X (i.e., {0, 2, 4, 6, 7, ...}), mientras que *u-suc*(X) generaría sólo una serie cuyo primer elemento es el *menor límite superior* de X (i.e., {6, 7, 8, ...}, quedando excluidos 0, 2 y 4). Sin

finita, verificamos primero $P(x_0)$ y a continuación verificamos $P(x_{n+1})$, habiendo supuesto previamente $P(x_n)$ para un número natural n cualquiera. Si X forma una serie transfinita, verificamos entonces $P(x_\alpha)$ (para un ordinal, α, cualquiera) partiendo del supuesto de que $P(x_\beta)$ se cumple para todo $\beta < \alpha$. Las pruebas por inducción e inducción transfinita son posibles gracias a On. En el caso de las definiciones por recursión ordinal, tenemos un conjunto X cuyos elementos se definen recursivamente mediante un operador iterativo, h (un generador del tipo aquí estudiado), que define cada elemento de X a partir de los elementos que ha definido previamente según la ecuación: $x_\alpha = h(\{x_\beta: \beta < \alpha\})$.[13]

La importancia de On y la relación que existe entre los ordinales, la generación de series y la propia concepción de (buen) orden, se ponen de manifiesto en uno de los axiomas más relevantes de cualquier teoría de conjuntos (incluya o no conjuntos no fundamentados): el *Axioma de Elección*. Según este axioma, dado un conjunto cualquiera X, siempre es posible elegir *primero* uno de sus miembros, *después* otro y así sucesivamente. Una manera precisa de formular este axioma es decir que: "Para cualquier conjunto X, existe una *función de elección*, *f*, entre $\wp(X) - \{\varnothing\}$ y X que a cada subconjunto, Y (no vacío), de X le asigna un

embargo, *u-suc*(X) ilustra mejor que m.o.n(X) cómo se construyen los ordinales partiendo de $X = \varnothing$.

[13] El principio de recursión ordinal se puede formular del siguiente modo (véase K. Devlin 1993, pp. 51-6): Sea h: On × V → V una función (dado que On y V, la jerarquía acumulativa, son clases, h es una función-clase) y sea λ un ordinal, entonces existe una única función(-clase) f: $\lambda \to V$ tal que, para todo $\alpha \in \lambda$, $f(\alpha) = h(\alpha, f|\alpha)$. Es decir, $f(\alpha)$ es igual al valor que le asigna h al conjunto de todos los valores generados por la restricción de f a α ($f|\alpha$). Se puede ver que (para $\alpha < \lambda$): $x_\alpha = f(\alpha) = h(\alpha, f|\alpha) = h(\{x_\beta$:

elemento de $X \cap Y$". Esto es, la función *f elige* siempre un miembro de X que está en Y: $f(Y) \in Y$. Si identificamos $f(Y)$ ($Y \neq \varnothing$) con *el menor elemento* de $Y \subseteq X$, es fácil ver que una consecuencia inmediata de la existencia de esta función es la posibilidad de definir un *buen orden* en X, ya que todos sus subconjuntos (salvo $\varnothing$) tienen un primer elemento: $f(X)$ elige el *primer* elemento de X, $f(X - \{f(X)\})$, el *segundo*, etcétera. Hecha esta identificación, podemos formar una *secuencia* con los elementos de X mediante un generador, *g*, valiéndonos de *f*. Si $Y \subset X$ (i.e., si Y es un subconjunto propio de X), definimos $g(Y)$ del siguiente modo: $g(Y) = f(X - Y)$ (i.e., $g(Y)$ es "el menor elemento de $X - Y$"). Así pues, una formulación equivalente del Axioma de Elección establece que *todo* conjunto X puede ser bien ordenado en una serie o, lo que es lo mismo, que sus elementos pueden ser puestos en relación biunívoca con un ordinal (i.e., X es isomorfo a un ordinal).[14] Volveremos a estas cuestiones más adelante, los ordinales y los generadores serán importantes no sólo a la hora de interpretar el Esquema de Inclusión sino también a la hora de abordar cierto tipo de soluciones a algunas de las paradojas reflexivas más populares. Pero nos centraremos primero en el papel que EI juega en toda esta historia.

$\beta < \alpha\}$). En el fondo $x_\alpha \in X$ es el valor que le asigna una función al ordinal $\alpha < \lambda$, una función cuyo dominio es λ (isomorfo a X) y cuya imagen es X.

[14] Cualquier manual de teoría de conjuntos incluye discusiones sobre estos temas: orden, series, ordinales, axioma de elección. Uno de los más claros e ilustrativos precisamente en estas cuestiones es P. Halmos 1960. Otro manual recomendable y utilizado en este trabajo es K. Devlin 1993.

3.2.5 El Esquema de Inclusión, cómo interpretar δ: diagonalizadores

Vimos en el capítulo anterior que el Esquema de Inclusión contiene básicamente dos elementos: una totalidad Ω (cuyos miembros satisfacen cierta propiedad φ) y una función δ de $\wp(\Omega)$ a Ω (restringida a los subconjuntos de Ω que satisfacen una propiedad ψ, satisfecha por Ω). La función δ cumple, además, dos condiciones adicionales a las que denominamos antes *clausura*: $\delta(X) \in \Omega$; y *trascendencia*: $\delta(X) \notin X$. Cabe preguntarse ahora cuál es el lugar de las paradojas y de EI en el contexto del presente discurso. ¿Qué relevancia pueden tener las ideas de generador, serie, orden, totalidad absoluta o límite iterativo en la formación de las contradicciones de inclusión? La idea fundamental –que anticipo ya– es la siguiente: mediante los elementos de EI podemos interpretar las paradojas reflexivas como problemas relacionados con nuestra incapacidad para concebir coherentemente una totalidad absoluta Ω que agrupe a todas las cosas que satisfacen cierta propiedad φ. Las paradojas se producen siempre cuando nuestra representación de la propiedad φ y de la totalidad Ω de las cosas que la satisfacen ("los φs") involucra de algún modo la aplicación de δ, un tipo de generador al que llamaremos diagonalizador. El problema radica en la posibilidad de aplicar δ sobre Ω (su límite iterativo) para definir (circularmente) un miembro de Ω ($\delta(\Omega) \in \Omega$) que a la vez trasciende Ω ($\delta(\Omega) \notin \Omega$). En esta sección y en la siguiente trataremos de interpretar y justificar la contribución hecha por cada uno de los elementos de EI a la obtención de una contradicción paradójica. Comenzaremos por la función δ.

Si nos fijamos en δ, nos daremos cuenta de que es posible interpretar esta función como un generador transfinito. Siguiendo a Priest, llamaré "diagonalizador" a este tipo de función-generador δ.[15] Examinemos esta analogía de cerca. La función δ de $\wp(\Omega)$ a Ω puede considerarse un generador en la medida en que algunos subconjuntos X de Ω (aquellos que satisfacen la propiedad ψ) pueden ser tratados como secuencias y δ(X) como el elemento que las "*continúa*". Continuar la secuencia X significa aquí añadirle δ(X), obteniendo así una secuencia nueva, X $\cup$ {δ(X)}, que está incluida en Ω y que puede ser "continuada", a su vez, mediante δ. Estas secuencias satisfacen, obviamente, las restricciones impuestas sobre δ. Por ejemplo, supongamos que partimos de una serie singular incluida en Ω: $\{x_0\} \subseteq \Omega$. Si aplicamos δ sobre esta serie, obtendremos $\delta(\{x_0\})$. Como δ satisface la condición de *trascendencia*, sabemos que $\delta(\{x_0\}) \notin \{x_0\}$ y, por tanto, el resultado de incluir $\delta(\{x_0\})$ en $\{x_0\}$ nos proporcionará una nueva serie, $\{x_0, x_1\}$, con dos elementos *distintos*: x_0 y $x_1 = \delta(\{x_0\})$. Al satisfacer δ la condición de *clausura*, sabemos, además, que $x_1 = \delta(\{x_0\}) \in \Omega$ y, por tanto, $\{x_0, x_1\} \subseteq \Omega$, con lo que podemos aplicar δ ahora sobre $\{x_0, x_1\}$ para obtener un tercer elemento, x_2. Si eventualmente formamos de esta manera una serie infinita $\{x_0, x_1, x_2, ...\} \subseteq \Omega$ y aplicamos sobre ella δ de nuevo, obtendremos un elemento transfinito, x_ω, tal que $\{x_0, x_1, x_2, ..., x_\omega\} \subseteq \Omega$. Si repetimos la operación una y otra vez hasta alcanzar su límite iterativo,

[15] "A diagonaliser can be thought of as a transfinite generalisation of what I have so far called a generator. We can define a transfinite sequence of members of Ω, $<w_\alpha; \alpha \in$

obtendremos una serie $T \subseteq \Omega$ (que en algunos casos coincidirá con Ω) más allá de la cual no podemos aplicar δ.[16] T no es necesariamente el producto de un generador transfinito. T puede ser el producto de un ω-generador o incluso de un generador que alcanza su límite iterativo tras un número finito de pasos, todo depende de cuándo se agoten las posibilidades de iterar δ desde su punto de partida. Cuando lleguemos a un $X \subseteq \Omega$ tal que $\delta(X)$ no esté definido o nos lleve fuera de Ω, T habrá alcanzado su techo dentro de Ω. Es importante señalar que cualquier función de elección (i.e., generador) que defina un buen orden dentro de un conjunto X (i.e., una serie con todos los elementos de X) satisface la descripción de δ que acabamos de dar y puede interpretarse consiguientemente como un *diagonalizador*.[17] La razón por la que estas funciones-generadores, δ, reciben el nombre de "diagonalizadores" está relacionada con el modo en que operan: δ se aplica siempre sobre una

---

On> by repeated application of the diagonaliser (transfinite recursion), thus: $w_\alpha = \delta(\{w_\beta; \beta < \alpha\})$. (If $\alpha = 0$, the argument of the function is just $\varnothing$)." *BLT*, p. 130.

[16] Todos los elementos de la serie $T = \{x_0, x_1, x_2, ...\}$ han de ser generados por δ, esto afecta también a x_0. El primer elemento de T depende del *primer subconjunto* de Ω sobre el que aplicamos δ. Si ese subconjunto es X, el primer elemento de la serie es: $\delta(X) = x_0$; el segundo, $\delta(X \cup \{x_0\}) = x_1$; el tercero, $\delta(X \cup \{x_0\} \cup \{x_1\}) = x_2$; etcétera. Los elementos de X *no pueden pertenecer* a T, ya que no son generados por δ ni en su primera aplicación ni en las siguientes (X está incluido en todas los subconjuntos X' sobre los que opera δ y, debido a la condición de trascendencia, $\delta(X') \notin X'$). Cuando tomamos $\varnothing$ como punto de partida, *todos* los elementos de los subconjuntos-series sobre los que opera δ son generados por δ. Esto simplifica la descripción de T: $x_0 = \delta(\varnothing)$; $x_1 = \delta(\{x_0\})$; x_2 =, $\delta(\{x_0, x_1\})$, etcétera. Obviamente, hemos elegido $\varnothing$ (por simplicidad expositiva) como punto de partida en el ejemplo de la explicación anterior, pero esto sólo es posible cuando $\varnothing$ está en el dominio de δ, cosa que no siempre ocurre.

[17] Comprobaremos esta afirmación en el siguiente apartado, veremos que sólo $\delta(\Omega)$ representa una variación significativa con respecto a otros generadores que ordenan bien un conjunto.

secuencia ordenada $X \subseteq \Omega$ para definir, a partir de los miembros de X, un nuevo elemento, $\delta(X)$, de Ω (*clausura*) que no puede coincidir con ninguno de los miembros de X (*trascendencia*) so pena de contradicción. En las demostraciones por *diagonalización* suponemos (con vistas a una *reductio*) que cierta secuencia o totalidad Ω agota todas las cosas que son φ y suponemos también que el diagonalizador δ es un generador de "φs" definible para Ω. De este modo, al aplicar δ sobre Ω obtenemos una contradicción: $\delta(\Omega)$ pertenece (*clausura*) y no pertenece a Ω (*trascendencia*).

3.2.6 El Esquema de Inclusión, cómo interpretar Ω y ψ: el alcance de δ

Hemos visto en qué sentido cabe interpretar δ como un generador, sin embargo falta por ver con más detenimiento cuáles son, dentro de EI, los límites de las secuencias que genera. Varios factores son relevantes en este sentido. El primer límite obvio con el que tropieza δ es Ω. Por la condición de clausura, sabemos que δ sólo puede generar cosas que sean φ, así pues, Ω (la totalidad de "los φs") aparece como una barrera infranqueable. Ω especifica la naturaleza de las cosas que puede generar δ aportando una propiedad φ (ser una oración verdadera, o un conjunto fundamentado, o un ordinal definible, etc.) característica de todas ellas. Todas las series que forma δ íntegramente están incluidas en Ω.

Pero, además de Ω, existen otras limitaciones que nos ayudan a precisar qué series forma δ y cuál es su "techo" iterativo dentro de Ω. Las limitaciones más relevantes vienen determinadas por el *dominio* de aplicación de δ –sólo podemos generar series iterando δ sobre conjuntos que estén en su dominio de aplicación– y por el *punto de partida* escogido para iniciar la iteración –las series generadas serán diferentes según δ parta de uno u otro elemento de su dominio. El papel que desempeña ψ en el Esquema de Inclusión es, precisamente, el de fijar el dominio de δ restringiéndolo a todos aquellos subconjuntos de Ω que satisfagan la propiedad ψ. Este dominio nunca es completamente arbitrario, pues ψ ha de asegurar, al menos, el cumplimiento de una condición:

(1) δ debe estar definida para Ω (i.e., Ω es ψ).

En la medida en que δ se interprete como un generador finito, un ω-generador o un generador transfinito, ψ ha de asegurar, además, otra condición:

(2) Siempre que δ no haya llegado a su límite iterativo α (sea éste finito o no), ha de ocurrir que:

(2a) si δ está definido para $X \subseteq \Omega$, debe estarlo también para $X \cup \{\delta(X)\} \subseteq \Omega$ (i.e., si X es ψ, entonces $X \cup \{\delta(X)\}$ es ψ.)

(2b) si δ está definido para una serie infinita de conjuntos, X_0, X_1, ..., X_β, ... $\subseteq \Omega$ cuyo límite es λ, entonces también lo ha de estar para $\bigcup_{\beta<\lambda} X_\beta \subseteq \Omega$ (i.e., si X_0, X_1, ..., X_β, ... son ψ, entonces $\bigcup_{\beta<\lambda} X_\beta$ es ψ.)

(2) implica que, dadas dos series con elementos de Ω, por ejemplo, X = $\{x_\beta \in \Omega: \beta < \alpha\}$ y X' = $\{x_\beta \in \Omega: \beta < \lambda\}$ (donde α y λ son ordinales y $\alpha < \lambda$), si X' se puede obtener a partir de X mediante aplicaciones de δ, es decir, si X' es una *continuación* de X mediante δ,[18] no puede ocurrir que X sea ψ y X' no. Si existe δ(X), existe también δ(X').

Entre las interpretaciones de ψ que satisfacen (1) y (2) hallamos dos casos extremos: en una, ψ se convierte en una propiedad universal (por ejemplo, "ser idéntico a sí mismo"), incluyendo en el dominio de δ a *todos* los subconjuntos de Ω; en otra, ψ es la propiedad "ser idéntico a Ω", reduciendo dicho dominio a su mínima expresión. La interpretación que hagamos de ψ es relevante a la hora de explicar cómo encajan diferentes razonamientos en EI. Por ejemplo, si δ está definida para *todos* los subconjuntos de Ω (i.e., si ψ es una propiedad universal), entonces Ω puede generarse *íntegramente* a partir de $\varnothing \subseteq \Omega$ mediante iteraciones de δ. Por el axioma de elección (véase 3.2.4), sabemos que todo conjunto (y en particular Ω) puede ser bien ordenado. Ahora bien, partiendo de un buen orden, $\leq$, sobre Ω, podemos construir un generador que defina una serie con todos sus elementos (véase 3.2.3). El generador en cuestión es (para todo X $\subset \Omega$): g(X) = "el menor elemento (según $\leq$) de Ω – X". Cada x $\in \Omega$ responde a la ecuación: $x = g(\{y \in \Omega: y < x\})$. Es fácil mostrar que g es un diagonalizador, ya que: g(X) es siempre el menor elemento de Ω (*clausura*) que no pertenece a X (*trascendencia*). Así pues, g puede equipararse a la función δ que aparece en EI. La única disonancia es Ω. g está definido sólo para subconjuntos *propios* de Ω (puesto que no existe $g(\Omega)$ = "el menor elemento de $\Omega - \Omega$") mientras que δ, por contra, siempre está definida para Ω. Para salvar este escollo estipularemos que $\delta(\Omega) = \Omega$ –por razones que quedarán más claras en la próxima sección– y que $\delta(X) = g(X)$, para todo $X \subset \Omega$. Esta discrepancia no es de extrañar, sin embargo, ya que el problema de las contradicciones de inclusión gira siempre en torno a $\delta(\Omega)$, un elemento de Ω que se define por referencia a Ω mismo y que, a la vez, pertenece a Ω y lo trasciende. La aparente necesidad de considerar $\delta(\Omega)$ para generar Ω con *todos* sus elementos es, precisamente, el origen de las paradojas.[19]

Priest denomina "contradicciones en los límites de la iteración" a las paradojas que encajan en EI cuando ψ es una propiedad universal. La característica fundamental de estas contradicciones es que en ellas *todos* los elementos de Ω (los "φs") son generados por δ. El conjunto de los "φs" y el conjunto de las cosas generadas por δ *son el mismo*. El límite iterativo, T, de δ partiendo de $\varnothing$ es el propio Ω. Sin embargo, no todas las contradicciones de inclusión son iguales. Cuando ψ no es una propiedad universal, la iteración de δ no genera Ω sino sólo subconjuntos propios de Ω. En estos casos, el problema de las contradicciones de

[18] X' es una continuación de X si y sólo si $X \subseteq X'$ y X es un segmento inicial de X' (i.e., existe un $a \in X'$ tal que $X = X'_a$. Véase 3.2.4).

[19] Si aceptamos el axioma de elección, se puede demostrar que existe una función δ: $\wp(\Omega) \rightarrow \Omega \cup \{\Omega\}$, tal que $\delta(\Omega) = \Omega$ y $\delta(X) \in \Omega - X$ cuando $X \subset \Omega$ (véase Devlin 1993, pp. 58-59, lema 2.7.2). La existencia de la función-generador δ no supone, pues, ningún problema. El problema radica más bien en que las totalidades paradójicas Ω satisfacen $\Omega \cup \{\Omega\} = \Omega$, cosa que *no* ocurre con conjuntos fundamentados.

inclusión no se puede relacionar con la posibilidad de construir Ω iterando δ. Para que Ω no sea generado íntegramente por δ basta con que uno solo de sus elementos, $a \in \Omega$, quede excluido de todos los subconjuntos sobre los que opera δ. Si no existe ningún $X \subseteq \Omega$ tal que $\delta(X \cup \{a\})$ exista, entonces no existe ninguna serie, X, generada por δ tal que $\delta(X) = a$, en caso contrario, $\delta(X \cup \{a\})$ existiría por la segunda condición impuesta sobre ψ. Así pues, tenemos un $a \in \Omega$ no generado por δ. Como vemos, el límite *iterativo*, T, de δ (i.e., la serie más grande con elementos de Ω que δ puede generar) está condicionado por los subconjuntos de Ω sobre los que actúa (los "ψs") y por el subconjunto del que parte para formar la serie.[20] Desde esta perspectiva, ψ aparece como el elemento de EI que determina cuál es el alcance real de δ como generador, ψ determina si δ genera todo lo que hay en Ω o sólo una parte. Priest utiliza diferentes nombres para referirse a las contradicciones de inclusión que sólo encajan en EI cuando ψ no es una propiedad universal: "contradicciones en los límites de la concepción; de la expresión; o de la cognición". Más adelante nos ocuparemos de ellas (3.2.9-10).

[20] Un ejemplo ayudará: Consideremos el generador m.o.n antes descrito. Vimos que si m.o.n parte de $0 = \varnothing$, da origen a On, la serie completa de los ordinales; si m.o.n parte de $\{0\}$, en cambio, da lugar a la serie $\{1, 2, 3, ...\}$; si parte de ω a $\{\omega+1, \omega+2, ...\}$. Como m.o.n está definido para todos los subconjuntos de On, si m.o.n parte de $\{0, 3\}$ la serie que genera es $\{1, 2, 4, ...\}$. En resumen, si m.o.n parte de $X \subseteq On$, da origen a una serie cuyos elementos son On – X, ya que, por la condición de trascendencia, nada de lo que aparece en el conjunto X de partida será producido por m.o.n (o por cualquier otro diagonalizador).

3.2.7 Contradicciones en los límites de la iteración: Burali-Forti [7]

Examinaremos primero las "contradicciones en los límites de la iteración", aquellas cuya estructura describió Russell (1905b) –recordemos (2.4 y 2.5) que estas contradicciones satisfacen el Esquema de Russell *simpliciter*–. Acabamos de ver que, en todas ellas, ψ es una propiedad universal satisfecha por todos los subconjuntos de Ω y, por ello, irrelevante. El rasgo estructural esencial que justifica su inclusión en una misma subfamilia es que en ellas Ω se puede entender indistintamente como la totalidad de los "φs" o como la totalidad de lo generado por δ, ya que x es φ si y sólo si x es producido por δ. Las paradojas aparecen aquí asociadas a la posibilidad de concebir Ω como el resultado de iterar δ. El alcance del problema se pone de manifiesto cuando nos percatamos de que afecta incluso a nuestra concepción más elaborada de lo que significa "formar una serie". Tropezamos con paradojas incluso al pensar en On, la "serie-patrón" que construimos antes para determinar "cómo construir una serie". Abordaremos este ejemplo con la intención de ilustrar otras contradicciones en los "límites de la iteración".

Si identificamos φ con la propiedad de ser un ordinal y convertimos a Ω en On, obtenemos una ejemplificación de EI. En este caso, m.o.n (el generador de On que vimos en 3.2.4) opera como un diagonalizador: m.o.n(X) asigna a cualquier X ⊆ On el menor ordinal (*clausura*) que no está en X (*trascendencia*), es decir, "el menor elemento (según ⊆) de On – X". m.o.n se puede entender aquí como una relación que asigna a cada subconjunto de On un único elemento de On, por lo que, mediante el Axioma de Reemplazo [*Replacement Axiom*], podemos establecer que m.o.n es una función de $\wp$(On) a On. El único elemento problemático del dominio de m.o.n es On, sin embargo, como veremos en breve, también a On le asigna m.o.n (paradójicamente) un ordinal: m.o.n(On) = On.

Además de la función m.o.n, necesitamos establecer también la existencia de On (ψ resulta aquí irrelevante por ser una propiedad universal) para tener una ejemplificación del Esquema de Inclusión. Las intuiciones que tenemos al respecto se apoyan en los axiomas de la teoría de conjuntos. Todos los ordinales se pueden construir apelando a los axiomas de Existencia, Infinitud y Reemplazo y a las operaciones conjuntistas (unión, etc.) reguladas, a su vez, por los axiomas correspondientes. Así pues (aunque revisaremos esta conclusión en el próximo capítulo), si todo lo que pertenece a On existe y está bien formado, On también debería existir. On es, de hecho, la unión de todos los ordinales, el límite iterativo de m.o.n. Tenemos, pues, una ejemplificación de EI:

1) Ω = On = $\{x$: x es un ordinal$\}$ existe (y, trivialmente, On = On).

2) Si X $\subseteq$ On (y, trivialmente, X = X), m.o.n(X) $\in$ On y m.o.n(X) $\notin$ X.

Veíamos en el capítulo anterior que cuando todas estas condiciones se satisfacen tropezamos con un problema. El problema se llama aquí paradoja de Burali-Forti [7] y surge cuando nos preguntamos si On es un ordinal. Recordemos (véase 3.2.4) que X es un ordinal si y sólo si (X, $\subseteq$) es un conjunto bien ordenado tal que, si $a \in X$, entonces $a = X_a$. De aquí podíamos inferir que todo ordinal α es igual al conjunto de sus predecesores estrictos según la relación de orden '$\subseteq$', $\alpha = \{\beta: \beta \subset \alpha\}$. Decíamos también que todo ordinal α es un subconjunto de On ($\alpha \subseteq$ On) y que $\bigcup\alpha = \alpha$ y $\cap\alpha = \varnothing$. Es fácil demostrar que (On, $\subseteq$) satisface la definición de ordinal (si $\alpha \in On \rightarrow \alpha = On_{\alpha}$) y cumple todas las propiedades antes mencionadas, por tanto, nos vemos forzados a concluir que On es un ordinal, el mayor de todos. On es, de hecho, un ordinal límite: la unión de todos los ordinales. Ahora bien, sabemos que todo ordinal α tiene un sucesor, $\alpha' = \alpha \cup \{\alpha\}$, distinto de él: $\alpha \neq \alpha'$.[21] Si reunimos todos estos hechos, tenemos una contradicción: Por una parte, On' = On $\cup$ {On} debe ser un ordinal distinto de On y mayor que él (On $\subseteq$ On'); por otra, al contener On a todos los ordinales, On' ha de pertenecer a On y, por tanto, ser menor que él (On' $\subseteq$ On), con lo que tenemos una contradicción: On = On', puesto que *ningún ordinal es igual a su sucesor*. En el lenguaje del Esquema de Inclusión, tenemos que m.o.n(On) = On' y podemos demostrar, por las condiciones de clausura y trascendencia, que m.o.n(On) $\in$ On (i.e., m.o.n(On) es un ordinal sucesor) y que m.o.n(On) $\notin$ On (i.e., m.o.n(On) $\in$ On se rechaza por *reductio* al implicar una falsedad: m.o.n(On) = On).

Resumiendo: m.o.n es un generador que construye la serie de los ordinales, On, formando en cada estadio un ordinal nuevo a partir de los ordinales formados en estadios precedentes. On satisface todas las propiedades que caracterizan a los ordinales y se puede concebir, de

[21] $\alpha \neq \alpha'$porque podemos demostrar por inducción transfinita que $\alpha \in \alpha'$ pero $\alpha \notin \alpha$.

hecho, como un ordinal límite: el resultado de aplicar m.o.n sobre la colección de todos los ordinales. Ahora bien, esto quiere decir que podemos obtener un ordinal *nuevo*, On $\cup$ {On}, aplicando m.o.n a On, lo cual es paradójico porque, al ser un ordinal, m.o.n(On) debe estar ya en On, el límite iterativo de m.o.n.

Otra paradoja que encaja en EI y cuya explicación desde el citado esquema se corresponde punto por punto con la de Burali-Forti [7], es la paradoja de Mirimanoff [9].[22] El Esquema de Inclusión nos ofrece una manera iluminadora de entender cuál es el problema que presentan ambas. Estas paradojas son, básicamente, un fenómeno relacionado con

---

[22] Consúltese [9] en el apéndice final. En este caso, Ω es R, el conjunto de todos los conjuntos fundamentados, que coincide con la *unión* de los elementos de la secuencia transfinita: $\langle x_0 = \varnothing, x_1 = \wp(\varnothing), x_2 = \wp(\wp(\varnothing)), ..., x_\omega = \bigcup\{x_n: n < \omega\}, ...\rangle$. Cada miembro de esta secuencia satisface la ecuación: $x_\alpha = \bigcup\{\wp(x_\beta): \beta < \alpha\}$ (donde $x_\alpha \subseteq$ R), así pues, R se puede entender como el límite iterativo del operador transfinito de la nota 5: *u-sub*($\{x_\beta: \beta < \alpha\}$). Podemos demostrar que *u-sub* es un diagonalizador:

(1) *u-sub* se puede generalizar a una función, ***u-sub***: $\wp$(R) $\rightarrow$ R, tal que ***u-sub***(X) = $\bigcup\{\wp(y): y \in$ X};

(2) ***u-sub*** satisface las condiciones de *clausura* y *trascendencia*: ***u-sub***(X) $\in$ R y ***u-sub***(X) $\notin$ X

[***Clausura***: *Si **u-sub**(X) no estuviese fundamentado, algún elemento de X tampoco lo estaría, contradiciendo X $\subseteq$ R. Por tanto, X está fundamentado*: ***u-sub***(X) $\in$ R]

[***Trascendencia***: *Si **u-sub**(X) $\in$ X, entonces todos los elementos de $\wp$(**u-sub**(X))* pertenecen a ***u-sub***(X), ya que ***u-sub***(X) = $\bigcup\{\wp(y): y \in$ X}. *Ahora bien, **u-sub**(X) $\in$ $\wp$(**u-sub**(X)) (i.e., todo conjunto es un subconjunto de sí mismo), por tanto*: ***u-sub***(X) $\in$ ***u-sub***(X) = $\bigcup\{\wp(y): y \in$ X}. *Esto significa que **u-sub**(X) no es un conjunto fundamentado, lo que contradice la propiedad de clausura **u-sub**(X) $\in$ R. Así pues, inferimos **u-sub**(X) $\notin$ X*].

Por último, al ser R la unión de todos los elementos de la secuencia descrita, se cumple que R = $\bigcup\{\wp(y): y \in$ R} y, por tanto, R = ***u-sub***(R). Esto presenta a R como un elemento más de la secuencia, lo que establece una contradicción de inclusión: ***u-sub***(R) $\notin$ R (*trascendencia*) y ***u-sub***(R) $\in$ R (*clausura*).

nuestra capacidad de pensar en los límites de aplicación de un operador iterativo, *g*, o, equivalentemente, con la idea de concebir como un "todo definido" el dominio absoluto de lo generado mediante *g*. La idea es la siguiente: Tomemos δ, una función que se puede interpretar de modo natural como un generador transfinito. Este operador genera siempre (en cada paso y sobre la base de lo anteriormente generado) algo del "mismo tipo", algo que satisface invariablemente una propiedad determinada φ (*clausura*): ser un ordinal, ser un conjunto fundamentado, etc.; y que, además, es nuevo, no ha sido generado previamente por δ (*trascendencia*). Cuando pensamos en lo que podemos construir mediante δ, de algún modo intentamos agrupar conceptualmente a todos los "φs" en una colección y pensamos en ella como una totalidad definida: la *totalidad absoluta*, Ω, de todos los "φs" obtenibles mediante δ. Más allá de Ω no cabe iterar δ, Ω define, pues, el *límite* de aplicación de δ. Sin embargo, cuando reflexionamos sobre cómo se construye Ω, nos damos cuenta de que todo lo que contiene "es φ" y se ha generado mediante δ. Así pues, y dado que δ es un operador transfinito que actúa sobre conjuntos de "φs" generados por δ, Ω se puede entender como el *resultado* de aplicar δ sobre *todos* sus elementos, es decir, como *el límite de la serie completa de los* "φs" (ordinales, conjuntos no fundamentados, etc.). El problema radica en que, desde esta perspectiva, Ω aparece como un "φ" más (esto es, como un elemento "límite" generado por δ) y, consiguientemente, nada nos impide continuar la serie iterando δ a partir de Ω. Esto produciría forzosamente *nuevos* "φs" (*trascendencia*), pero eso es imposible, porque todos los "φs" deben estar ya en Ω (*clausura*).

Hablar de paradojas en los límites de la iteración es hablar de las paradojas que genera hoy en día la noción de infinito. En todas estas paradojas identificamos métodos para generar por iteración ciertas totalidades Ω (los números ordinales, los números cardinales, los conjuntos bien fundamentados, etc.) Una vez generadas estas totalidades, nos damos cuenta de que tienen la misma naturaleza que sus elementos y que nada nos impide aplicar los mecanismos de formación de ordinales, cardinales, etc. sobre Ω para prolongar la serie con nuevos elementos (en particular, con Ω). Sin embargo, la serie era, por definición, la más grande que podíamos concebir en su género, con lo que tenemos una paradoja.

3.2.8 ¿Contradicciones en los límites de la iteración?: Russell [2]

La explicación que acabamos de dar al discutir las paradojas de Burali-Forti [7] y Mirimanoff [9] es la explicación que Priest parece asociar en general con las "contradicciones en los límites de la iteración". Además de [7] y [9], Priest incluye en este grupo a las paradojas de Russell [2] y Cantor [10].[23] Es un hecho que todas ellas satisfacen el

[23] Priest sitúa aquí también lo que él denomina "quinta antinomia" [19] (*BLT* pp. 100-01, 130-31), basada en las antinomias kantianas. A simple vista se aprecia que [19] (véase apéndice final) satisface la explicación dada para [7] y [9] perfectamente. En este caso, Ω es P y ψ una propiedad universal. El generador p(x) se puede convertir fácilmente en una función, δ, de $\wp$(P) a P, que satisface las propiedades de clausura (p(x) pertenece a P) y trascendencia (p(x) no es idéntico ni a x ni a ninguno de sus elementos: p(x) $\notin$ x). Así pues, p(x) forma una secuencia cuyo límite absoluto es P,

Esquema de Russell *simpliciter*, siendo ψ una propiedad universal irrelevante. Sin embargo, encontramos aquí una discrepancia importante. Hemos sugerido implícitamente que todas las contradicciones de inclusión requieren *interpretar* δ en el sentido en que lo hace Priest, i.e., como un generador (transfinito) que origina una secuencia definiendo cada uno de sus miembros a partir de todos sus antecesores: $\delta(\{x_\beta\colon \beta < \alpha\}) = x_\alpha$. Aunque esto es cierto en casi todos los casos, no siempre se cumple, hay excepciones relevantes, sobre todo una: la paradoja de Russell [2]. En 2.4 constatamos que [2] encaja en EI si identificamos δ con la función *identidad*. Vimos que, si $\Omega = \{x\colon x \notin x\}$ (φ = "no pertenecerse a sí mismo" y ψ = "ser idéntico a sí mismo"), la identidad podría interpretarse como una función de $\wp(\Omega)$ a Ω (véase la nota 16 del capítulo 2) que satisface las condiciones de trascendencia y clausura. Sin embargo, está claro que la identidad *no es* en ningún sentido un generador del tipo descrito por Priest, 'id(x)' *no define x a partir de sus predecesores en una secuencia*. Obviamente, si δ ha de ser la identidad, entonces son necesarios *todos* los subconjuntos de Ω para "generar" Ω, pero ahí se acaba el paralelismo con paradojas como [7] y [9]. La propiedad de clausura viene garantizada en [2] por el mero hecho de que δ "repite" lo que ya encontramos en Ω, y la de trascendencia porque, según φ, los miembros de Ω no se pertenecen a sí mismos, así pues, δ no explica en ningún momento *cómo se genera* Ω. Estas diferencias con respecto a [7] o [9] (o [19]) son relevantes y tienen, además, un carácter

---

ahora bien, nada nos impide pensar en P, lo que implica la existencia de p(P), un *estructural.* Del mismo modo que reducir ψ a una propiedad universal parece convertirla en un factor irrelevante en la descripción de ciertas contradicciones de inclusión, reducir δ a la función identidad parece disminuir su peso en [2]. Priest pasa por alto este punto, pero esto no le resta relevancia. Recordemos que es el propio Russell (1905b) quien defiende originalmente el paralelismo entre [2], [7], [9] y [10] y quien diseña un patrón estructural en el que todas estas paradojas tienen cabida (identificando a δ con la función identidad en [2]). Priest fundamentalmente amplia el esquema russelliano y ofrece una interpretación relevante de sus ingredientes esenciales, pero es justo señalar que su interpretación chirría en este punto. Esto no representa ningún problema para el Esquema de Inclusión, pero debe forzarnos a revisar la etiqueta "contradicciones en los límites de la iteración" en relación con paradojas como [2].

Hay, al menos, un sentido en el que el peso de δ en EI disminuye al ser identificada con la función identidad. Cuando esto ocurre, no hay en principio nada que nos indique cómo se genera Ω, cómo formar una secuencia con sus miembros mediante un operador δ. Podemos interpretar este hecho como una señal clara de que la contradicción es independiente del modo en que se ha generado Ω. La contradicción se sigue aquí de la propia naturaleza de Ω: de la(s) propiedad(es) φ que caracteriza(n) a sus elementos y de la relación que Ω mantiene consigo mismo. En estas circunstancias, llamar a la paradoja de Russell [2] una

---

miembro límite de la secuencia, pero entonces tenemos $p(P) \in P$ y $p(P) \notin P$.

contradicción "en los límites de la *iteración*" no resulta adecuado, ya que no se apela a la "iteración" de operador alguno, [2] no es el fruto de imaginar el límite iterativo de un generador que va formando una colección de cosas que son φ. Existe, no obstante, cierta relación entre todas las contradicciones en las que ψ es una propiedad universal, sean contradicciones en los límites de la iteración o no (i.e., sea δ un generador o la función identidad). En la medida en que Ω puede ser bien ordenado –como dicta el axioma de elección (3.2.4)– y δ puede estar definido para todos sus subconjuntos, sabemos (3.2.3 y 3.2.6) que es posible construir un generador interpretable como una función de $\wp(\Omega)$ a Ω y equivalente a un buen orden sobre Ω. Esto significa que, en el caso de la paradoja de Russell, por ejemplo, es posible construir una función δ de $\wp(\Omega)$ a Ω, *distinta de la función identidad*, que genera todos los elementos de Ω (la clase de todos los conjuntos que no se pertenecen a sí mismos) y que desencadena una contradicción de inclusión en los límites de la iteración para el valor $\delta(\Omega) = \Omega$. Este resultado no es de extrañar, ya que muchas de las paradojas en las que ψ es una propiedad universal tienen por objeto *la misma* colección Ω y sólo difieren en el generador empleado para formarla (si emplean alguno), o en la propiedad φ a través de la cual caracterizamos a los elementos de Ω. Este es el caso (en teorías clásicas de conjuntos, como ZF) de las paradojas de Russell [2], Mirimanoff [9] y Cantor [10]. La clase de "todos los conjuntos que no se pertenecen a sí mismos", la de "todos los conjuntos fundamentados", la de "todos los conjuntos generados mediante ***u-sub***" o, simplemente, la de "todos los conjuntos", son la misma clase en cualquier teoría que excluya conjuntos no fundamentados. Pese a este hecho, hay que insistir en que la paradoja de Russell [2] *no* es una contradicción en los límites de la iteración, sencillamente involucra una colección Ω susceptible de ser generada íntegramente mediante un diagonalizador y, por tanto, de provocar contradicciones en los límites de la iteración.

3.2.9 Contradicciones en las que ψ no es una propiedad universal

El problema central en todas las contradicciones de inclusión gira siempre en torno al mismo punto: la legitimidad de aplicar δ sobre Ω. Siempre que sea posible definir $\delta(\Omega)$, inferiremos una contradicción. Si, pese a ello, encontramos aún razones sólidas para aceptar $\delta(\Omega)$, tendremos además una paradoja. Lo que varía en cada caso es el camino que nos conduce a $\delta(\Omega)$ y las razones que lo justifican.[24] En las paradojas examinadas en este capítulo (excluyendo Russell [2] y Cantor [10]), la clase Ω era un producto de la iteración de δ que, al estar definido para todos sus subconjuntos, nos permitía generar u ordenar bien Ω.[25] Esto en sí no representa un problema, ya que la inmensa mayoría de generadores y conjuntos bien ordenados *no* son paradójicos. Sin embargo, la peculiaridad del conjunto Ω en las contradicciones iterativas radica en que, a diferencia de lo que ocurre con otros conjuntos, la mayor serie generada por δ (i.e., Ω) incluye forzosamente a $\delta(\Omega)$. Pero, ¿qué razones

---

[24] Con la salvedad de la paradoja de Russell [2] y afines, donde la presencia de δ es meramente testimonial.

teníamos para aceptar $\delta(\Omega)$? Básicamente tres: (1) Teníamos motivos para pensar que las propiedades "ser φ" y "ser generado por δ" tenían la misma extensión. Una inducción transfinita (o, a veces, una mera inspección de la definición de Ω) mostraba que todos los "φs" eran generados por δ y viceversa. (2) Había razones para pensar que Ω existía y satisfacía la propiedad φ. Dado que "ser φ" es condición necesaria y suficiente para pertenecer a Ω y para ser generado por δ, inferíamos dos hechos: a) $\Omega \in \Omega$ y b) existe un $X \subseteq \Omega$ tal que $\delta(X) = \Omega$. Por último, (3) había razones para pensar que Ω era el resultado de iterar δ sobre Ω, ya que parecía intuitivo concebir Ω como un elemento *límite* de la mayor secuencia transfinita generada por δ: "el resultado de aplicar δ sobre el conjunto de *todos* los 'φs' generados por δ (i.e., sobre Ω mismo)": $\delta(\{x_\alpha: \varphi(x_\alpha) \wedge \alpha \subset On\}) = \delta(\Omega) = \Omega$.[26]

El problema de estas contradicciones radica, pues, en que Ω es la mayor serie posible generada por δ, pero, pese a ello, δ no se detiene al llegar a Ω, su límite iterativo, sino que lo transgrede continuando la serie con $\delta(\Omega)$. Es pertinente preguntarnos ahora qué ocurre cuando tropezamos con una contradicción de inclusión en la que ψ no es una propiedad universal y, por consiguiente, δ no está definida para todos los subconjuntos de Ω. En estos casos δ también es un diagonalizador y Ω un

[25] De hecho, cualquier subconjunto $X \subseteq \Omega$ se puede generar u ordenar bien mediante δ partiendo de $\delta(\Omega - X)$. Ya vimos que Ω se generaba a partir de $\delta(\varnothing) = \delta(\Omega - \Omega)$.

[26] Como la operación iterativa δ es irrelevante en el caso de la paradoja de Russell [2], sobran las razones (1) y (3). Omitiendo toda referencia a δ, la paradoja se justifica a través de (2): Hay razones para pensar que Ω satisface φ, esto es, que Ω es un conjunto

conjunto bien ordenado (como todos), sin embargo, δ no puede ser el operador que genera u ordena bien Ω. Ciertamente, δ sólo genera cosas que son φ pero no puede generar a *todos* "los φs". Como vimos en 3.2.6, δ puede generar, a lo sumo, algunos subconjuntos *propios* de Ω. Esto significa que ninguna de las series que forma δ es igual a Ω y, por tanto, que el problema de estas contradicciones, y de Ω en particular, no reside en las posibles series iterativas que forma δ dentro de Ω sino exclusivamente en la posibilidad de definir o construir δ(Ω). Pero, ¿qué motivos podemos tener para aceptar δ(Ω) en aquellos casos en los que ψ *no es* una propiedad universal? La respuesta depende de cómo interpretemos ψ dentro de EI.

La misión de ψ es describir aquellas condiciones que, al ser satisfechas por un $X \subseteq \Omega$, justifican la construcción de δ(X). ψ selecciona aquellos subconjuntos de Ω sobre los que puede actuar δ. En este sentido, cuanto más sencilla y fácil de satisfacer para Ω sea la propiedad ψ, más visos tendrá una contradicción de inclusión de convertirse en una paradoja.[27] En la mayoría de las contradicciones que ahora nos ocupan, ψ es la propiedad "ser definible (o descriptible) bajo *x* parámetros" (i.e.: "en un lenguaje *L* con tales características", "en

---

que no se pertenece a sí mismo, de donde se sigue, por ser φ condición necesaria y suficiente de pertenencia a Ω, que Ω pertenece a Ω, una contradicción.

[27] En la "paradoja" del barbero [18] (véase la nota 30 del capítulo 2), Ω era el conjunto de todas las personas y ψ la propiedad "ser un grupo de personas en el que aquellos que no se afeitan a sí mismos son afeitados por una única persona". Es obvio que Ω no satisface ψ, por eso [18] es sólo una pseudo-paradoja. No tenemos razón alguna para pensar que exista un barbero que afeite a todas las personas que no se afeitan a sí

castellano y con menos de n sílabas", etcétera). Cuando ψ no es ninguna de estas propiedades, es habitualmente una propiedad satisfecha por Ω y tal que, si X satisface ψ, se cumple siempre que X es definible o descriptible bajo x parámetros.[28] (También cuando ψ es la propiedad "ser igual a Ω" asumimos implícitamente que Ω es un conjunto definible o descriptible.) La insistencia en la "definibilidad" o "descriptibilidad" bajo ciertos parámetros de los argumentos de δ, obedece a nuestra voluntad de asegurar que $\delta(X)$, y sobre todo $\delta(\Omega)$, existen y están definidos bajo las condiciones estipuladas por el enunciado del problema. Dado que la operación δ parece existir y ser definible –sabemos "en teoría" cómo opera sobre sus argumentos, aunque sea imposible computar a veces todas sus aplicaciones–, nos falta asegurar que sus argumentos (o, al menos, Ω) existen y son definibles o descriptibles en un sentido que facilite la acción de δ sobre ellos para producir algo que "sea φ". Si nos resultara del todo imposible referirnos al conjunto Ω, describirlo, definirlo o pensar en él, la expresión '$\delta(\Omega)$' carecería de sentido, no se podría calcular, definir, describir o construir, puesto que no podríamos conocer Ω, ni siquiera pensar en su existencia. En las contradicciones en

---

mismas, podemos rechazar la supuesta contradicción porque no tenemos *razones* para aceptar $\psi(\Omega)$.

[28] Recordemos que "ser definible" supone aquí para Priest poco más que "ser descriptible de algún modo" (véase la nota 18 del capítulo 2). En algunas de las paradojas que considera (no tratadas en este trabajo), ψ es una propiedad del tipo "ser pensable", "ser imaginable", o "ser concebible". Se entiende aquí que si algo es pensable, imaginable o concebible, puede ser representado o definido (en el sentido de ser descrito) de algún modo. La paradoja de Berkeley (*BLT*, pp. 60-70, 132-135) es un ejemplo paradigmático. También se supone la transición de "ser pensable" a "ser descriptible" en la "quinta antinomia" [19].

los límites de la iteración, no teníamos que preocuparnos por la definibilidad o existencia de los conjuntos X sobre los que actuaba δ porque *todos* ellos eran construidos por δ. Partiendo de $\varnothing$, δ generaba la serie Ω y, sobre la base de ese buen orden, generaba cualquier subconjunto $X \subseteq \Omega$ con $\delta(\Omega - X)$ como primer elemento. Definir δ suponía implícitamente explicar cómo se podría definir o describir, al menos "idealmente", cualquier conjunto de su dominio, incluyendo Ω.[29] En las contradicciones que ahora nos ocupan, en cambio, definir o describir δ no garantiza por sí mismo que todos los subconjuntos de Ω sean definibles o descriptibles del modo que requiere la contradicción de inclusión, ya que la existencia de Ω no depende de la acción de δ exclusivamente. Los subconjuntos de Ω que no son fruto de la iteración de δ –entre ellos Ω– no pueden definirse apelando a δ y por ello no podemos dar por sentada sin más su definibilidad. Esto es lo que pone de manifiesto ψ en EI, la interpretación de ψ (cuando no es una propiedad

[29] "Idealmente" quiere decir aquí "en un lenguaje con suficientes recursos expresivos". Si intentásemos, por ejemplo, nombrar o definir en nuestro lenguaje a *todos* los ordinales, nos faltarían nombres o descripciones (la paradoja de König [5] explota esta circunstancia). Por su cardinalidad, ningún lenguaje natural puede asociar un término distinto a cada ordinal, ya que sus expresiones, a diferencia de los ordinales, son enumerables. Otro tanto se puede decir con respecto a la posibilidad de describir *cada* aplicación de m.o.n, la función que genera On. Sabemos cómo opera m.o.n "en general", pero no podemos describir todas sus operaciones. Ahora bien, esto no significa que haya ordinales o aplicaciones de m.o.n "esencialmente" indescriptibles. Una cosa es afirmar que *carecemos* de una descripción o de un lenguaje adecuado para describir algo y otra bien distinta afirmar que ese algo sea indescriptible en todo lenguaje posible. Si existiesen lenguajes de cardinalidad idéntica o superior a la de On, podríamos describir cada ordinal α mediante expresiones del tipo: $\alpha = \text{m.o.n}(\{\beta: \beta < \alpha\})$. Otras veces, el problema es distinto. A veces asumimos que nombres o descripciones de nuestro lenguaje describen "algo" que, sin embargo, no existe (un "círculo cuadrado") y, por tanto, no puede "ser descrito".

universal) es siempre lo que asegura directa o indirectamente que Ω y, por consiguiente, $\delta(\Omega)$ sean definibles.

Hay en principio algo que nos empuja a creer que el conjunto Ω que aparece en las contradicciones de inclusión existe y se puede definir o describir. Ese "algo" es la posibilidad de identificar Ω con una "totalidad absoluta", la extensión de cierta propiedad φ, y de pensar en ella bajo una definición o descripción definida del tipo: "El conjunto de todos los x que satisfacen la propiedad φ". Si esto es así, es fácil que $\psi(\Omega)$ se cumpla cuando interpretamos ψ como "ser definible", "descriptible", "imaginable", etc. porque siempre podremos definir, describir o imaginar Ω como "el conjunto de todos los φs". Ahora bien, ¿es esto realmente posible? En principio nos sentimos inclinados a pensar que sí. Desde el momento en que reflexionamos sobre cierta propiedad φ, tiene sentido preguntarse "qué objetos satisfacen dicha propiedad", "cuál es el número de todos los φs", "qué otras propiedades comparten todos los φs o en qué se diferencian unos de otros". Todas estas preguntas presuponen de algún modo la existencia de una totalidad definida (que podría ser vacía, como ocurre con "el conjunto de todos los x que son círculos cuadrados"). Pero esta intuición es precisamente una de las cosas que ponen en tela de juicio las paradojas que estamos examinando en este trabajo. Después de todo, parece razonable concluir que determinado conjunto o totalidad no puede existir si de su supuesta existencia se siguen absurdos o contradicciones. La mayoría de soluciones clásicas a las paradojas reflexivas recurren a estrategias de este tipo para bloquear la inferencia de una contradicción. Como veremos en el próximo capítulo, Priest no simpatiza con ninguna de estas soluciones y utiliza, en cambio, varios argumentos para justificar la existencia de este tipo de totalidades. Sin embargo, nos ocuparemos de estas cuestiones en su momento, cuando abordemos el problema de las posibles soluciones a las paradojas reflexivas.

Entre las paradojas que sólo encajan en el Esquema de Inclusión cuando ψ no es una propiedad universal encontramos las siguientes: el mentiroso [1] (en todas sus variantes), Berry [4], König [5], Richard [6], Grelling [8], la cita [11], Barwise-Moss [12], Yablo [13], el hiper-juego [15], el conocedor [*knower's paradox* 16], Curry [17] y también la paradoja relacional [3] entre muchas otras. Algunas fueron tratadas en el capítulo anterior. Podemos encontrar una reducción al Esquema de Inclusión de todas estas paradojas (y del resto con la excepción de [11], [12] y [15]) en Priest 1994 y *BLT* ([13] está en Priest 1997a y [3] está en *BLT* p. 143 nota 2). Remitimos al lector a estas obras para ver cómo encajan en EI las paradojas citadas que no hemos tratado aquí.

3.2.10 Contradicciones en las que ψ sólo es satisfecha por Ω

Dentro de las contradicciones en las que ψ *no es* una propiedad universal el alcance de ψ es de nuevo relevante para introducir ulteriores distinciones. Las "contradicciones en los límites de la iteración" (aquellas en las que ψ *es* una propiedad universal) representan un caso límite dentro de EI. Las contradicciones en las que ψ es una propiedad *singular* satisfecha sólo por Ω representan el caso límite opuesto. Estas contradicciones se ajustan a lo ya expuesto en el apartado anterior. Priest las denomina "contradicciones en los límites de la expresión" debido a que todas parecen mostrar una cierta unidad temática. Todas ellas aparecen en teorías del significado cuya tesis principal es que ciertos hechos son "inefables", teorías que, sin embargo, se autorrefutan al conseguir, paradójicamente, hablar sobre lo que no se puede hablar.[30] Priest no explica en ningún momento cuál es la relación que existe entre esta unidad temática y el hecho de que la extensión de ψ se reduzca a Ω en estas contradicciones. Por otra parte, todos los ejemplos que aporta tienen una marcada carga teórica, aparecen en el seno de complejas teorías sobre la relación del lenguaje con el mundo. El único ejemplo sencillo (además del teorema de Cantor en 2.7) de contradicción que encaja en esta subfamilia es la paradoja de la cita de Tarski [11] tratada en 2.6. El compromiso teórico de esta contradicción de inclusión es mínimo, por eso resulta más verosímil su condición de paradoja genuina frente a los ejemplos discutidos por Priest. Para desembocar en una contradicción, [11] sólo requiere interpretar las comillas como funciones que asignan a cada término lingüístico un nombre del mismo, lo que, en

[30] Priest lo expresa de este modo: "All of the theories of meaning that we have looked at in this part of the book [cuarta parte de *BLT*] render some important states of affairs ineffable (whilst managing to express them). And in all cases the states of affairs in question are ones that are, in some sense, about notions that make expression in language itself possible: the unity of the proposition (Frege, earlier Wittgenstein), reference (Quine), truth (Davidson [...]), rule-following (later Wittgenstein), *différance* (Derrida). We have here in all cases an inclosure contradiction [...] $\varphi(y)$ is 'y can be expressed in language' (where y is some state of affairs), so that Ω is the totality of things that can be expressed; $\psi(x)$ is '$x = \Omega$'; $\delta(\Omega)$ is s, where this is the relevant state of affairs concerning whatever notion it is that makes expression in language possible. Then by the argument in each case we have $\neg\varphi(s)$ (Transcendence); yet in the process of analysis we demonstrate that $\varphi(s)$ (Closure)." *BLT*, p. 223. Véase también pp. 192 y

principio, parece natural e intuitivo. También esta paradoja "acepta" la etiqueta de "contradicción en los límites de la expresión", ya que lo que pone de manifiesto son ciertas limitaciones relativas a nuestra capacidad de "nombrar" objetos coherentemente. Pese a todo, esta etiqueta no refleja ningún rasgo estructural y es dudoso que se aplique a todas las contradicciones de inclusión de este tipo. No se aplica, por ejemplo, al teorema de Cantor.

3.2.11 Otros criterios estructurales para clasificar las contradicciones de inclusión

Si nos fijamos en el papel que desempeña ψ en EI, nos daremos cuenta de que cobra especial relevancia cuando intentamos demostrar que el operador δ satisface la condición de *clausura*. $\psi(X)$ garantiza que $\delta(X)$ sea definible, descriptible, expresable o demostrable dentro de los parámetros que aseguran la pertenencia de $\delta(X)$ a Ω. ψ sólo se convierte en una propiedad irrelevante cuando la propia naturaleza del operador δ y del conjunto Ω escogidos aseguran que todo lo que produce δ será φ y que $\delta(X)$ estará siempre bien definido. Cuando estas condiciones no se pueden asegurar apelando a δ y Ω, necesitamos que ψ garantice la condición de clausura. Hasta ahora nuestras clasificaciones se han centrado en las interpretaciones que ψ recibía en EI. Sin embargo ψ no es el único elemento de EI que justifica distinciones estructurales entre las contradicciones de inclusión. Otros rasgos formales nos permiten trazar

146-7. La cuarta parte de *BLT* está dedicada a las "contradicciones en los límites de la expresión".

distinciones y generar subfamilias dentro de estas contradicciones con independencia de cómo interpretemos ψ en EI. Estos rasgos se centran sobre todo en los diferentes modos que tenemos de demostrar la condición de *trascendencia* para el operador δ.

Aunque δ es siempre un diagonalizador y las demostraciones de trascendencia adoptan invariablemente la forma de una *reductio*, las razones por las que llegamos a $\delta(X) \notin X$ a partir de la suposición de $\delta(X) \in X$ varían ligeramente. Hay ciertas contradicciones de inclusión (Berry [4], König [5] o Richard [6], por ejemplo) en las que δ es un operador descriptivo, un operador que selecciona o describe un miembro de Ω a partir de un conjunto $X \subseteq \Omega$. En estas contradicciones $\delta(X) \notin X$ se puede demostrar *directamente* a partir del hecho (fácilmente contrastable) de que existen elementos de Ω que no pertenecen a X: $\exists x\ (x \in \Omega - X)$. Todo lo que tenemos que hacer para asegurar que $\delta(X) \notin X$ es hacer que $\delta(X)$ describa a uno de esos elementos. Tomemos como ejemplo la paradoja de König [5], donde Ω es "el conjunto de todos los ordinales definibles" y δ el operador "el menor ordinal que no está en ..." (es decir, m.o.n). En esta paradoja, $\delta(\Omega)$ es m.o.n(Ω), "el menor número ordinal indefinible". Se puede demostrar que si $X \subseteq \Omega$ (i.e., si X es un conjunto de ordinales definibles), entonces existen ordinales que no pertenecen a X, ya que $\Omega \subset$ On. Uno de esos ordinales es "el menor elemento de On – X", cuya existencia viene garantizada por el hecho de que On está bien ordenado. Ahora bien, la expresión 'm.o.n(X)' describe precisamente a ese ordinal, con lo que establecemos, de forma directa, la condición de trascendencia para δ: m.o.n(X) $\notin$ X. (Así mismo, al satisfacer X $\subseteq \Omega$ la propiedad ψ de "ser definible", tenemos que m.o.n(X) es *definible*, razón por la que m.o.n(X) $\in \Omega$ obteniendo así la condición de clausura.) Priest se refiere a todas estas contradicciones con el nombre genérico de "contradicciones en los límites de lo definible (de lo descriptible o de lo concebible)". Priest afirma además que, en las demostraciones de clausura y trascendencia, todas las paradojas que subsume esta categoría comparten una misma estructura formal a otros niveles.[31]

Existe, en segundo lugar, un grupo de paradojas (el mentiroso [1], Grelling [8], la cita [11], Yablo [13] o el conocedor [16], por ejemplo) en las que $\delta(X) \notin X$ se prueba sólo de forma *indirecta*. Para encontrar un elemento en Ω que no pertenezca a X e identificarlo con $\delta(X)$ necesitamos el concurso de un *principio* semántico o epistémico que justifique la validez de la implicación: $\delta(X) \in X \rightarrow \delta(X) \notin X$. Priest denomina a estos principios "principios puente" porque nos hablan de cómo se relacionan nuestro lenguaje y nuestro conocimiento con el mundo. El Esquema-T tarskiano es uno de los "principios puente" más frecuentes en estas paradojas: $x \in T \leftrightarrow p$ (versión "conjuntista" de 'x es verdadero $\leftrightarrow p$', donde T es el conjunto de todas las oraciones –o proposiciones, etc.– verdaderas, 'p' una oración y 'x' un nombre de 'p' –o de lo expresado por 'p'). En el caso del mentiroso [1], por ejemplo, Ω es T, el conjunto de "todas las oraciones (o proposiciones, etc.) verdaderas" y δ un operador que a cada X $\subseteq$ T le asigna una oración

[31] Véase *BLT*, pp. 131-33. También lo que Priest denomina "paradoja de Berkeley" encaja en esta categoría.

(proposición, etc.) μ = '$\mu \notin X$'. (La paradoja del mentiroso es $\delta(T)$ = '$\mu \notin T$'.) A partir de $\delta(X) = \mu$ no podemos inferir *directamente* $\delta(X) \notin X$, para ello necesitamos el Esquema-T. De este modo, al suponer $\delta(X) \in X$ (i.e., $\mu \in X$), establecemos primero $\mu \in T$ (puesto que $X \subseteq T$) y después, $\mu \in T \leftrightarrow \mu \notin X$ (por el Esquema-T y μ = '$\mu \notin X$'), de donde inferimos finalmente $\mu \notin X$ (i.e., $\delta(X) \notin X$). Este tipo de paradojas es bautizado en *BLT* con el nombre de "contradicciones en los límites de la cognición".[32]

Priest parece presentar las diferencias estructurales que hemos discutido en las pruebas de "trascendencia" para δ, como rasgos genuinos de dos grupos de paradojas: las contradicciones en los límites de la definibilidad y en los límites de la cognición. Se considera que ambos grupos son independientes de las contradicciones en los límites de la iteración y de la expresión. Cabe decir al respecto que, si Priest tiene razón y los cuatro grupos son independientes entre sí, ha de ser sobre la base de rasgos estructurales adicionales compartidos por todos los miembros de cada grupo. Las pruebas directas que caracterizan a las contradicciones en los límites de la definición aparecen también en algunas contradicciones en los límites de la iteración (Burali-Forti [7] o Mirimanoff [9]), donde ψ es una propiedad universal. Mientras que las pruebas indirectas con principios puente (características de las contradicciones en los límites de la cognición) aparecen también en algunas contradicciones en los límites de la expresión (donde ψ es una propiedad cuya extensión es el conjunto singular $\{\Omega\}$). Éste es el caso de la paradoja de la cita [11], que hace uso de dos principios puente. Uno

[32] Esta distinción se encuentra en *BLT*, pp. 141-46.

nos permite pasar de *'x' es demostrable* a *x* (*'x' es demostrable* → *x*), el otro nos permite pasar de una serie finita de oraciones '*a*, *b*, *c*, ..., *x*' que satisfacen ciertas condiciones a la afirmación '*x es demostrable*'. Por este motivo, la clasificación que introduce ψ en las contradicciones de inclusión parece independiente de la que introducen los criterios ahora examinados. Lo que no impide que la clasificación de Priest sea en última instancia correcta sobre la base de la coincidencia de varios rasgos estructurales (no sólo de uno o dos) en cada uno de los grupos que esboza.

Por último, cabe señalar que a veces tropezamos con divergencias estructurales aparentes que en el fondo resultan espurias. Ya discutimos en el capítulo anterior (2.8.3) la más llamativa: Al formular algunas paradojas que encajan en EI (Russell [2] o Burali-Forti [7]) mencionamos una *totalidad* ("los conjuntos que no se pertenecen a sí mismos", "los ordinales"). Sin embargo, al formular paradojas como la del mentiroso, que también encajan en EI, no mencionamos ninguna totalidad, sencillamente encontramos una oración que se atribuye la propiedad de "no ser verdadera". Priest apunta que esta diferencia es en el fondo aparente, ya que existe una fuerte conexión entre las nociones de "satisfacción de una propiedad φ" y la de "pertenencia al conjunto que forma la extensión de φ". Desde este punto de vista, todas las paradojas que prescinden de totalidades en su formulación y hablan, en cambio, de "satisfacción de la propiedad φ" se pueden reformular hablando de "pertenencia a la extensión de φ (es decir, a $\Omega = \{x\text{: } \varphi(x)\}$)". 'Esta oración no es verdadera' pasaría a ser 'Esta oración no pertenece al conjunto de lo verdadero'. Del mismo modo, las paradojas que hablan de totalidades se pueden parafrasear en términos de la satisfacción de ciertas propiedades (el propio esquema EI puede ser parafraseado en este sentido, véase 2.8.3, nota 31).[33]

3.2.12 ¿Cuál es el problema de fondo en las contradicciones de inclusión?

Es importante recalcar que el problema de las contradicciones de inclusión radica *exclusivamente* en $\delta(\Omega)$. La noción de "diagonalizador" o la idea de "totalidad cuyos miembros se definen mediante cierta propiedad φ" no son problemáticas en ningún sentido, ni siquiera lo son cuando se cumple que la totalidad de "los φs" coincide exactamente con el conjunto de todo lo generado mediante un diagonalizador. De hecho, hablar de "generadores", "funciones", etc. es un mero formalismo, cualquier función se puede concebir "en ciertas circunstancias" como un generador de "cierto tipo" y viceversa. En este sentido, la noción de generador no es en sí misma culpable de las contradicciones de inclusión, como tampoco lo es la de diagonalizador. Básicamente, un diagonalizador δ es un generador (i.e., un operador iterativo) que actúa sobre (conjuntos de) cosas de tipo φ produciendo siempre cosas nuevas (*trascendencia*) de ese mismo tipo (*clausura*). Un sinfín de operadores triviales y muy respetables responde a esta descripción. Muchas son las funciones que, dependiendo del contexto y de las circunstancias, podrían

[33] Para la defensa general de esta idea, véase Priest 1987 (capítulos 2 y 10) y *BLT* (pp. 144, 279-80).

interpretarse como diagonalizadores.[34] Tampoco considerar "la totalidad, Ω, de los φs" representa un problema forzosamente. Ni "la totalidad, ω, de los números naturales" ni "la totalidad, *Seis*, de los números naturales definibles con seis dígitos en notación decimal" son totalidades contradictorias o paradójicas. Tanto para ω como para *Seis* podemos encontrar diagonalizadores que generen a todos sus miembros. Esta afirmación es sencillamente una constatación de que (siguiendo el Axioma de Elección) todo conjunto –y en particular ω o *Seis*– puede ser bien ordenado y de que existe un estrecho vínculo entre las nociones de "buen orden" y "serie generada por un diagonalizador". ω, por ejemplo, es generado (o bien ordenado) por ***suc*** y *Seis* por ***suc-6*** (un diagonalizador que opera igual que ***suc*** hasta llegar a ***suc-6***(999.999), a partir de ese momento deja de estar definido). *Seis* y ω son los límites iterativos de ***suc-6*** y ***suc*** respectivamente, cumpliéndose, tanto para *Seis* como para ω, que "ser φ" es lo mismo que "ser generado por δ" (i.e., por

[34] Cualquier función que se pueda definir por recursión ordinal es un generador (véase 3.2.4, en especial, nota 13). Por otra parte, la mayoría de generadores pueden ser interpretados en ciertos contextos como diagonalizadores. Hemos visto ya unos cuantos: *suc*, *tsuc*, *u-sub*, pero hay muchos más. Por ejemplo, la función constante *c* que a todo *x* le asigna el mismo valor: $c(x) = k$, se puede interpretar en ciertos contextos como el generador de la serie <*k*, *k*, ...>. Aquí *c* no podría ser un diagonalizador, ya que, a partir de su segunda iteración, *k* pertenece al conjunto de lo previamente generado mediante *c* (no se satisface la condición de trascendencia). En otro contexto, en cambio, *c* podría ser el diagonalizador (finito) que genera la secuencia singular <*k*>. Otro tanto se puede decir de la identidad con respecto a las series <*x*, *x*, ...> y <*x*> cuando id(x) tiene por dominio $\{x\}$. Lo mismo ocurre con otras funciones recursivas primitivas simples como las proyecciones en general. La función sucesor, *suc*, es siempre interpretable como un ω-generador y es, además, un diagonalizador. Las recursivas complejas (bien por composición, bien por el principio de recursión) también pueden ser entendidas como generadores en ciertos contextos.

suc-6 o ***suc*** respectivamente).[35] Lo que separa a ω y a *Seis* (*Seis* es igual al número natural 1.000.000) de las totalidades Ω que aparecen en las contradicciones de inclusión es que ***suc***(ω) y ***suc-6***(1.000.000) *no están definidas*, mientras que δ(Ω) sí lo está en las contradicciones de inclusión. ω y 1.000.000 *no están* en los dominios respectivos de ***suc*** y ***suc-6***, ni siquiera existe un x tal que ***suc***(x) = ω o ***suc-6***(x) = 1.000.000. El problema de las contradicciones de inclusión radica siempre en aquellos factores relativos a la naturaleza de Ω y de δ que nos imponen la validez de δ(Ω). Las contradicciones en los límites de la iteración no son problemáticas porque el conjunto Ω sea generado por un diagonalizador δ sino porque, para generar Ω de forma satisfactoria, debemos considerar como uno de sus elementos a δ(Ω). Cuando Ω no aparece en el dominio de δ, no hay contradicción de inclusión: Si δ genera todos los elementos de Ω, tenemos simplemente un buen orden sobre Ω y si no los genera, una operación inocua que define ciertos elementos de Ω a partir de otros.

3.2.13 Un problema filosófico

El objetivo de Priest en *BLT* no es meramente estudiar cuál es la estructura de las paradojas de autorreferencia o reflexivas. Más allá de este fin, lo que Priest persigue es mostrar *por qué* es relevante ese estudio. Una de las muchas virtudes de *BLT* es recordarnos que lo que hace de las paradojas un objeto digno de estudio son precisamente las conexiones que existen entre éstas y problemas filosóficos de hondo

[35] *Seis* también podría ser generado por ***suc***, aunque entonces "ser φ" y "ser generado por ***suc***" no coincidirían.

calado histórico. Una paradoja, una contradicción forzosamente derivada a partir de los principios que rigen nuestra concepción del infinito, de la semántica, de la epistemología o incluso de lo que podemos imaginar o concebir, es una muestra de que algo falla en nuestra imagen cotidiana del mundo. Habitualmente asumimos que las contradicciones no pueden tener validez o realidad objetiva y, sin embargo, como muestran las paradojas, concedemos validez objetiva a muchos principios de los que parece inevitable derivar una contradicción. Así pues, en algo nos debemos equivocar. ¿Por qué, entonces olvidamos ese hecho? El motivo esencial parece ser que, en la mayoría de casos, las paradojas se asocian a ideas, objetos u oraciones triviales que no tienen implicaciones prácticas relevantes en nuestra vida cotidiana. El mentiroso es quizá el mejor ejemplo. Con todo, no hay que olvidar (como nos advertía Tarski en la nota 4 del capítulo 1 o Ramsey en la cita de 2.2) que las paradojas introducen problemas recalcitrantes cuya solución resulta urgente en ciertos ámbitos teóricos, sobre todo en el campo de las ciencias formales. Y es precisamente en estos campos donde apreciamos el auténtico poder de las paradojas, un gran número de teoremas, cómo por ejemplo el que afirma la no existencia del conjunto de todos los conjuntos, son el fruto de reacciones tajantes a la paradoja de Russell [2].

Pero, aunque les prestemos menos atención, las paradojas también hacen acto de presencia en parcelas del lenguaje y ámbitos teóricos que no están formalizados. Las paradojas ocupan un lugar prominente, por ejemplo, dentro de la filosofía, donde resulta prácticamente imposible ignorarlas (lo que, como veremos, no es casual). En el segundo capítulo de este trabajo, defendimos, junto a Russell y Priest, que un buen número de paradojas y razonamientos comparten una misma estructura, la que refleja el Esquema de Inclusión. Desde este punto de vista, existe un nivel de abstracción desde el que todos ellos son ejemplos de un mismo problema general, lo que no impide que haya también diferencias relevantes entre ellas. Ahora bien, ¿cuál es el problema general?

Ésta es la pregunta que hemos intentado responder en el presente capítulo a través de un análisis de los elementos que componen EI y de su interpretación en cada caso. Toda contradicción de inclusión contiene dos elementos fundamentales, Ω y δ, y presenta un problema común, la aparente necesidad de considerar $\delta(\Omega)$. Vimos que Ω era una totalidad absoluta, el conjunto de *todas* las cosas que satisfacen cierta propiedad φ; y vimos también que δ era un operador que, a partir de un conjunto X de cosas que eran φ (y pertenecían, por tanto, a Ω), generaba siempre cosas que pertenecían a Ω (*clausura*), pero no a X (*trascendencia*). El problema surgía cuando, por diferentes razones (y es aquí donde cobran relevancia las diferencias estructurales entre contradicciones), nos resultaba imposible negar la existencia de $\delta(\Omega)$, por un lado, y su condición de "ser φ", por otro. Esto nos forzaba a pensar que $\delta(\Omega)$ era un elemento de Ω, lo que comporta, de entrada, cierto grado de circularidad incómodo, puesto que uno de los elementos de Ω, $\delta(\Omega)$, sólo se puede definir por referencia a Ω y viceversa. Pero el problema más grave es que, junto con $\delta(\Omega) \in \Omega$, hemos de aceptar también $\delta(\Omega) \notin \Omega$, por ser δ un diagonalizador y satisfacer la condición de trascendencia, obteniendo así una contradicción. En este contexto, Ω (la *totalidad* de "los φs") aparece como el límite iterativo de δ, y de cualquier otro operador que genere "φs", y δ aparece como un operador que transgrede ese límite absoluto. Esto hace que sea imposible concebir coherentemente a Ω en su totalidad y a δ en todas sus aplicaciones.

Si nos fijamos en la historia de la filosofía, nos percataremos al instante de que sus problemas giran a menudo en torno a conceptos que aluden a totalidades absolutas definibles a partir de operaciones iterativas que intentamos ejecutar mentalmente hasta alcanzar su límite absoluto, i.e., hasta alcanzar aquella totalidad más allá de la cual no cabe preguntarse si hay algo que pueda ser generado por dicha operación. Con frecuencia nos sorprendemos al descubrir que, en cierto modo, la idea de la existencia de un límite que no pueden traspasar ciertos operadores convive con la idea opuesta de que dicho límite es, por su naturaleza, el tipo de cosa sobre la que podríamos aplicar el operador en cuestión para encontrar algo nuevo. Ambas ideas parecen tener sentido y, sin embargo, se contradicen. A veces la situación es incluso más chocante. En ocasiones, cuando intentamos definir cuál es el límite de cierto operador u operadores, nos damos cuenta de que nuestra propia empresa está condicionada por esos límites, i.e., todo lo que podemos establecer al respecto depende de, o está generado directamente por, los operadores en cuestión. En este caso, cobramos conciencia de que cualquier afirmación positiva sobre dichos límites tendría, al menos, que vislumbrar lo que hay más allá de ellos, cosa imposible puesto que todo lo que podemos establecer al respecto ha de estar forzosamente *dentro* de esos límites. Este tipo de conceptos, ideas y situaciones abundan en la historia de la filosofía. La idea de átomo, por ejemplo, se corresponde con la idea de lo absolutamente indivisible; la de Dios, con la de lo absolutamente incondicionado (perfecto, omnisciente, todopoderoso, etc.); la de sujeto, marca los límites absolutos de la unidad de nuestra experiencia; cierta idea de "lenguaje" o "realidad lingüística" (no la de este o aquel lenguaje particular) marca los límites absolutos de lo expresable; la de conocimiento, los límites absolutos de lo cognoscible; la de razón los límites de lo inteligible; la de pensamiento, los de lo pensable, etcétera. Cuando hablamos de entidades "límite" de este tipo o intentamos fundamentarlas, bordeamos los límites de la coherencia, de lo que se puede decir sin caer en contradicciones.

Uno de los principales cometidos de Priest en *BLT* es examinar discusiones filosóficas históricas relacionadas con conceptos como los que acabamos de mencionar. Priest defiende que en ellas encontramos argumentos con la forma de contradicciones de inclusión, sin embargo, no todos ellos son paradojas (genuinas). Recordemos que, cuando es posible rechazar la existencia de la totalidad Ω o del operador δ o de $\delta(\Omega)$ sobre la base de razones sólidas, aunque el razonamiento que examinamos tenga la forma de una contradicción de inclusión, no es una paradoja (del mismo modo que un razonamiento con premisas falsas podría tener la forma de una aplicación correcta del *modus ponens*). Algunos de los argumentos que discute *BLT* cronológicamente se presentaron como paradojas, otros como una *reductio*. Priest aporta algunas de las razones históricas (y otras de cosecha propia) por las que se pensó que las "paradojas" en unos casos o la supuesta "*reductio*" en otros eran contradicciones inaceptables o argumentos erróneos fruto de teorías equivocadas. Priest se detiene, sin embargo, en una serie de contradicciones de inclusión que son firmes candidatos a ser tratadas como paradojas genuinas, contradicciones que aún no se han resuelto teóricamente de forma satisfactoria. Entre ellas encontramos incluso algunas (como Burali-Forti [7] o Russell [1]) cuya solución clásica en teoría de conjuntos satisface a muchos (aunque, obviamente, no a Priest). Estas paradojas constituyen el núcleo central de *BLT* y de su descripción y análisis de EI.

Tanto en su análisis histórico de las contradicciones de inclusión [*inclosures*] (que comprende las partes 1ª, 2ª, 4ª y 5ª de *BLT*) como en su análisis formal de EI (3ª parte), Priest clasifica a éstas bajo los rótulos de contradicciones en los límites de: (1) "la iteración"; (2) "la concepción"; (3) "la cognición"; y (4) "la expresión". La clasificación de Priest se debe en parte a criterios formales y estructurales, como ya vimos, pero éstos no parecen suficientes para identificar mediante condiciones necesarias y suficientes la pertenencia de una contradicción a uno u otro grupo. Por eso cabe decir que, al menos a este nivel, la clasificación de Priest no es estrictamente formal y, como ocurriera con la de Ramsey, obedece a criterios de "contenido". Quizá el bloque de paradojas más claramente delimitado sea el que encaja en el Esquema original de Russell, lo que Priest denomina "contradicciones en los límites de la iteración". Pero, aun así, ya vimos que la propia paradoja de Russell [2] no parece particularmente relacionada con la iteración y mantiene discrepancias estructurales relevantes con respecto a sus "compañeras" de grupo.

Entre las contradicciones en los límites de la iteración, Priest examina discusiones históricas relacionadas con las ideas de generar una serie infinita por iteración y de establecer sus límites, mostrando cómo, a través de sucesivas imágenes erróneas de la noción de infinito, nos hemos ido acercando progresivamente a nuestra actual concepción del mismo (y a las paradojas que vimos en 3.2.7 y 3.2.8).[36] Del mismo modo, aborda las contradicciones en los límites de la expresión hablando del Cratilo de Platón, de las ideas de "Hyle" (o materia primera) y de cambio sustancial en Aristóteles (también de las teorías mencionadas en la cita de la nota 30). Al hablar de las contradicciones en los límites de la cognición habla del escepticismo pirrónico de Sexto Empírico y del escepticismo académico de Carneades, así como del relativismo de Protágoras y de las paradojas del mentiroso y el conocedor. Por último, al abordar las contradicciones en los límites de la concepción habla del argumento ontológico de San Anselmo y de lo que Priest da en llamar "paradoja de Berkeley" acerca de la posibilidad de concebir cosas que no son concebibles.[37]

---

[36] Priest discute (*BLT*, capítulo 2) las paradojas de Zenón sobre el movimiento; la idea aristotélica de la imposibilidad del infinito *en acto*; el argumento cosmológico de Santo Tomás de Aquino a favor de la existencia necesaria de una primera causa, Dios, para evitar un *regreso al infinito* de causas (un infinito en acto); y la idea leibniziana de la necesidad de Dios como *razón suficiente* (extramundana) de la existencia (mundana) de una serie infinita de fenómenos causalmente ordenados. En los capítulos 5-7 habla de las antinomias kantianas y de la noción hegeliana de infinito, antes de pasar en los capítulos 8 y 11 a hablar de Cantor y la concepción actual del infinito.

[37] El argumento ontológico parece, de hecho, un ejemplo de totalidad infinita generada por iteración y diagonalización: Pensemos en un conjunto de "perfecciones" X. Si podemos pensar en una "perfección", x, tal que $x \notin X$, es posible construir un nuevo conjunto que contenga más "perfecciones": $X' = X \cup \{x\}$. "El ser más perfecto que podamos imaginar", Dios, ha de ser el límite de esta operación iterativa (i.e., la reunión de *todas* las perfecciones). Si la existencia es una "perfección", entonces ha de

No podríamos discutir aquí en detalle cuál es la relación que las paradojas de inclusión guardan con problemas filosóficos históricos concretos relacionados con la posibilidad de explorar los límites del pensamiento y del lenguaje. Priest ha llevado a cabo esa labor de forma excelente en *BLT* y remitimos al lector a esta obra para encontrar una sólida defensa de una tesis que creemos correcta: "Muchos de los problemas que han preocupado a los filósofos a lo largo de la historia tienen la *forma* de contradicciones de inclusión y están por ello emparentados con las paradojas reflexivas". Aquí (y también en la introducción y en las conclusiones) sólo hemos podido ofrecer algunos indicios a favor de esta tesis. Nuestro objetivo fundamental en estas páginas ha sido estudiar el *mecanismo* que describe EI y que provoca la inferencia de contradicciones. Hemos intentado entender la contribución que cada elemento de EI hace a la inferencia de una contradicción interpretando el significado de cada uno de ellos en casos típicos y describiendo su comportamiento: Ω es una totalidad absoluta, δ un generador (de cierto tipo: un diagonalizador) de elementos de Ω y ψ un conjunto de condiciones que permiten aplicar δ sobre Ω para obtener una contradicción.

---

pertenecer al concepto de Dios, de lo contrario no sería la reunión de *todas* las perfecciones y podríamos, por diagonalización, imaginar un "ser" *más* perfecto. (Priest rechaza también el argumento en *BLT*, pp. 56-60.) Como se puede apreciar, muchas veces los problemas de iteración, expresión, concepción, etc. aparecen entrelazados en un mismo argumento, por eso las subfamilias de contradicciones de inclusión que introduce Priest no parecen tan nítidamente basadas en factores estructurales.

3.3 Algunas intuiciones sobre la estructura de las paradojas reflexivas y EI

Antes de concluir este capítulo, nos gustaría poner de manifiesto una serie de virtudes del Esquema de Inclusión en las que *BLT* no incide demasiado. Decíamos en el capítulo anterior que ninguno de los análisis estructurales existentes de las paradojas reflexivas ofrecía condiciones suficientes para identificar este tipo de paradojas, aunque determinados rasgos que tradicionalmente se han asociado a ellas sí parecían ofrecer condiciones necesarias. Uno de los más significativos es el que justifica la etiqueta de "paradojas de autorreferencia o reflexividad" otorgada por Russell. Ciertamente, todas estas paradojas parecen ir indisociablemente unidas a algún tipo de autorreferencia, reflexividad o circularidad. Es este hecho, en primera instancia, lo que justifica la intuición de que deberíamos abordarlas de forma conjunta. Por este motivo sería conveniente mostrar que EI puede hacer justicia a esta intuición (y a otras) y dar cuenta de los rasgos mencionados. Pero también sería conveniente dar cuenta dentro de EI de otras ideas formales y estructurales que se ha asociado tradicionalmente a este tipo de paradojas y a sus soluciones.

3.3.1 Autorreferencia, reflexividad, circularidad y EI

En contra de lo que pudiera parecer, dado su frecuente uso en relación con estas paradojas, el término "*autorreferencia*" dista mucho de ser claro y preciso en este contexto. Como se puede apreciar a simple vista, muchas de las paradojas examinadas (Russell [2] o Burali-Forti [7], por ejemplo) no hablan sobre nociones semánticas ni, por tanto, sobre la relación de referencia; y tampoco parecen involucrar, al menos directamente, objetos lingüísticos autorreferentes. Así pues, no es en absoluto evidente qué queremos decir al hablar de "paradojas de autorreferencia". Tampoco el término "*reflexividad*" resulta mucho más preciso, aunque sea quizá más flexible a la hora de captar algo que todas estas paradojas podrían compartir. Hablar de reflexividad en este contexto sugiere la presencia de entidades que mantienen cierto tipo de relación *consigo mismas* ("auto-referencia", "auto-pertenencia", "auto-aplicabilidad", etc.). Pero también este término resulta demasiado vago y poco revelador.

Si queremos de verdad identificar con precisión un rasgo estructural compartido por todas las paradojas de autorreferencia (una condición necesaria –aunque no suficiente– para pertenecer a este grupo) debemos acudir de nuevo a las observaciones de Russell (véase 2.2, notas 5 y 6). Según Russell, en todas estas paradojas encontramos una totalidad X entre cuyos miembros se cuenta a, un elemento de X (a veces, a = X) que *sólo* puede definirse, describirse o concebirse apelando al *conjunto* de todos los miembros de X y, en particular, a a mismo. Esto implica que, al pensar en los miembros de X, tropecemos siempre con uno, a, cuya definición "nos remite (o refiere) a" sí mismo. Cuando tratamos de "reconstruir mentalmente" X, nos percatamos de que nuestra "reconstrucción" tropieza con un elemento, a, que sólo podemos "construir" si disponemos ya de X con todos sus elementos, incluido a. De este modo, cualquier intento de definir o imaginar a a o a X "nos remite (o refiere)", en última instancia, a ellos mismos. Éste es el único papel descriptivo apropiado (y más bien "metafórico") que pueden jugar las nociones de "autorreferencia" o "reflexividad" dentro del grupo de paradojas que nos ocupan. Sería un error pensar que todas estas paradojas contienen alguna entidad directamente autorreferente o reflexiva. En un sinfín de ocasiones, como por ejemplo en las cadenas de mentirosos:

A B es verdadera. B A es falsa.

Las entidades problemáticas (como A y B en este caso) no mantienen ninguna relación *directa* consigo mismas (bien sea de referencia o de cualquier otro tipo), sino sólo a través de intermediarios.

Lo que Russell describe aludiendo a la totalidad X y a uno de sus elementos, *a*, es una situación de *circularidad viciosa* que, bajo los parámetros expuestos en el párrafo anterior, puede ser descrita en términos de "autorreferencia" o "reflexividad". Una de las virtudes del Esquema de Inclusión es que nos permite acomodar perfectamente la intuición russelliana y mostrar que el rasgo por él descrito es, ciertamente, común a todas las paradojas que aquí nos ocupan. La totalidad de la que habla Russell no es otra que Ω en EI y el elemento problemático es, obviamente, $\delta(\Omega)$. Sabemos que δ es un operador que nos permite describir o definir un elemento de Ω (*clausura*) a partir de *todos* los elementos de (algunos de) sus subconjuntos y sabemos también que $\delta(\Omega)$ existe y define un elemento de Ω. Así pues, $\delta(\Omega)$ es un elemento de Ω que *sólo* puede definirse, describirse o concebirse por referencia a *todos* los miembros de Ω ($\delta(\Omega)$ incluido), reproduciendo así la situación de circularidad viciosa antes descrita. Así mismo, a veces ocurre (en las contradicciones en los límites de la iteración) que $\Omega = \delta(\Omega)$, como observábamos en el párrafo anterior. Se puede decir, pues, que EI capta perfectamente la idea de que las paradojas reflexivas involucran problemas de circularidad viciosa.

3.3.2 Negación y diagonalización

Decíamos en 2.3 que la situación de circularidad descrita por Russell no explica por sí misma el fenómeno de las paradojas. Aunque este tipo de estructuras circulares se reproduce en todas las contradicciones de inclusión (siendo condición necesaria de las mismas), no explica por qué inferimos una contradicción en cada caso. De hecho, cuando Russell describe las paradojas de autorreferencia (nota 5, capítulo 2) no sólo habla de circularidad, apunta también que una de las peculiaridades de éstas es el hecho de que, a partir de todos los elementos de X, definimos un a que a la vez *pertenece* y *no pertenece* a X, con lo que a parece agrandar la "totalidad absoluta" X más allá de sus límites. Ahora bien, el razonamiento que nos lleva a inferir $a \notin X$ es *independiente* del razonamiento que nos lleva a inferir $a \in X$. Ésta es la razón por la que encontramos numerosas entidades "circulares" (la oración del "veraz": 'Esta oración es verdadera'; o el conjunto $c = \{c\}$) que no derivan en contradicciones. Son a menudo entidades que ni siquiera involucran ningún tipo de partícula negativa. Es esto lo que nos llevó a rechazar el diagnóstico del problema que efectuó Russell en su artículo de 1908, el Principio de Circularidad Viciosa (véase la nota 6 del anterior capítulo). Debemos encontrar algo que marque la diferencia entre un caso de circularidad viciosa en el que, por ejemplo, la negación no juega ningún papel relevante y las contradicciones de inclusión.

Una de las intuiciones más recurrentes en la bibliografía sobre las paradojas reflexivas es la idea de que existe un estrecho vínculo entre éstas y cierto tipo de razonamientos a los que denominamos argumentos "diagonales". Básicamente, un argumento diagonal típico es un razonamiento que parte de dos premisas fundamentales:

(1) la existencia de una clase X que contiene a todas las cosas que satisfacen cierta propiedad P, i.e., X = $\{a, b, c, d, ...\}$ = $\{x$: P('x')$\}$

(2) la existencia de cierta manera de ordenar totalmente X mediante los miembros de una clase ya ordenada Y = $\{\alpha, \beta, \chi, \delta, ...\}$ con (supuestamente) un número de elementos igual o superior al de X.

En la práctica, la premisa (2) estipula la existencia de una función biyectiva entre X e Y (o un subconjunto de Y) que determina un ordenamiento completo de los miembros de X, i.e., X = $\{a_\alpha, b_\beta, c_\chi, d_\delta, ...\}$ (a menudo Y es el conjunto de los números naturales).

A partir de estas premisas, el argumento diagonal demuestra la posibilidad de construir, apelando a *todos* los miembros de X (y sirviéndose de la posición que cada uno de ellos ocupa en el orden que introduce Y y en las características de dicho orden), un z que satisface la propiedad P y que pertenece, por tanto, a X (i.e., $z \in$ X), pero que no puede aparecer en dicho ordenamiento ($z_{¿?}$), lo que contradice el supuesto de que el ordenamiento en cuestión afecta a *todos* los miembros de X, incluido z. Bajo la supuesta verdad de sus premisas, el argumento diagonal establece, pues, una contradicción: $z \in$ X (ya que P('z') se cumple) y $z \notin$ X (ya que no existe un $x \in$ Y tal que z_x, pese a existir una biyección entre X e Y). El más conocido ejemplo de argumento diagonal fue ofrecido por Cantor (véase 2.7) al demostrar que el conjunto de los números reales tiene una cardinalidad mayor que el de los números naturales –i.e., los números reales no pueden ser ordenados mediante los naturales suponiendo una relación biyectiva entre ambos–. El argumento se puede generalizar para establecer que el conjunto de todos los subconjuntos de un conjunto dado X tiene siempre una cardinalidad superior a la de X.[38]

En un sentido más amplio, también se utiliza el nombre de argumento diagonal para designar razonamientos que, partiendo de una premisa fundamental: la existencia o definibilidad de una totalidad X = {x: P('x')} cuyos miembros satisfacen cierta propiedad P, demuestran la posibilidad de construir un a que no puede pertenecer a X: $a \notin$ X, pero que cumple P('a') y nos permite inferir, por tanto, $a \in$ X. Al igual que en el caso anterior, el elemento problemático, a, se define o describe apelando a *todos* los elementos de X (aunque no necesariamente a su ordenamiento). El argumento diagonal más conocido (en este sentido amplio del término) es posiblemente el teorema de Tarski, según el cual es posible construir en ciertos lenguajes una oración μ (i.e., la oración del

[38] No entraremos aquí en detalles. El lector podrá encontrar ejemplos de argumentos diagonales (incluido el de Cantor), una definición sencilla y una buena muestra de la relación existente entre estos argumentos y las paradojas en K. Simmons 1993, capítulos 2-4.

mentiroso) tal que μ pertenece al conjunto de todas las verdades si, y sólo si, no pertenece a dicho conjunto (i.e., μ es verdadera si, y sólo si, no es verdadera).

A modo de resumen podemos decir que, dada la existencia de X –i.e., la clase de todas las cosas que satisfacen cierta propiedad P– (y, en algunos casos, de cierto orden total de los elementos de X), un argumento diagonal es un razonamiento que construye apelando a X (y, si procede, a su ordenamiento) un a tal que $a \in$ X y $a \notin$ X. Esta contradicción se emplea habitualmente en una demostración por *reductio* de la inexistencia o indefinibilidad de la totalidad X (o en una demostración por *reductio* de la imposibilidad de ordenar totalmente X del modo supuesto en las premisas del argumento). El hecho de que en la construcción de a (el elemento problemático de X) apelemos a la propia totalidad X, a *todos* sus elementos, muestra que los argumentos diagonales también incorporan un problema de circularidad.

La descripción de argumento diagonal que hemos ofrecido pone de manifiesto que su estructura es en buena medida la estructura que describe el Esquema de Inclusión, no en vano la función δ que en él aparece es bautizada por Priest con el nombre de diagonalizador. Así pues, también en este sentido EI hace justicia a las intuiciones que tenemos sobre las paradojas. En el Esquema de Inclusión suponemos la existencia de una totalidad, Ω, cuyos elementos satisfacen cierta propiedad φ. Así mismo, definimos un operador, δ, que, actuando sobre *todos* los elementos de un conjunto X de "φs", construye "un nuevo φ", i.e., un elemento de Ω (*clausura*) que no está en X (*trascendencia*). En la medida en que Ω es "un conjunto de φs", podemos aplicar δ sobre la propia totalidad Ω para obtener algo, $\delta(\Omega)$, que ha de ser un elemento de Ω pero que, a la vez, no puede serlo. Está claro que esta descripción encaja perfectamente con la descripción más amplia que hemos aportado de argumento diagonal.

Algunas ejemplificaciones de EI también encajan en la descripción (más restringida) de argumento diagonal que apela al ordenamiento de ciertas totalidades. Las contradicciones en los límites de la iteración: Burali-Forti [7], Mirimanoff [9], etc. pueden ser interpretadas fácilmente en este sentido. Podemos considerar aquí dos clases que teóricamente coinciden: La clase X de todas las cosas que satisfacen la propiedad φ (i.e., "ser un ordinal", "ser un conjunto fundamentado", etc.) y la clase Y de todas las cosas generadas mediante el operador δ (i.e., *u-suc*, *u-sub*, etc.). La hipótesis de partida es que ambas clases tienen el mismo número de elementos (de hecho, se supone que son *la misma*: X = Y = Ω) y que podemos, por tanto, ordenar los elementos de X mediante los elementos ya ordenados de Y (el orden de Y viene determinado por la secuencia s (s = Y) de elementos que genera δ por iteración partiendo de $\varnothing$). Desde esta perspectiva, $\delta(\Omega)$ nos ofrece un elemento de X = Y = Ω (i.e., "un φ") que no puede pertenecer a X y, además, lo hace apelando a todos los elementos de X (i.e., de Ω) y al modo en que estos han sido ordenados mediante un generador transfinito.

Pero las contradicciones en los límites de la iteración no son las únicas que admiten ser descritas en términos de argumentos diagonales que involucran la posibilidad de ordenar conjuntos de un determinado modo. La paradoja de Richard [6], que encaja en EI y es una contradicción en los límites de la expresión, también admite esta lectura, y, si cabe, de un modo más natural que las contradicciones en los límites de la iteración. Parece obvio, pues, que el Esquema de Inclusión refleja a grandes rasgos la estructura de un argumento diagonal. De hecho, un estudio estructural de diferentes estrategias de "diagonalización" podría enriquecer nuestro análisis estructural de las contradicciones de inclusión y de sus subfamilias. Del mismo modo, un estudio detallado de diversas estrategias para generar fenómenos de circularidad viciosa asociados a teorías de la referencia, conjuntos, etc. también haría progresar nuestra descripción de estas contradicciones.

El papel que juega la negación en las contradicciones de inclusión está asociado precisamente a δ, el operador que utilizamos para superar los límites de la totalidad Ω. La posibilidad de obtener una contradicción depende de las propiedades de clausura y trascendencia satisfechas por este operador. La propiedad de trascendencia (i.e., $\delta(X) \notin X$) es, en particular, la que introduce la negación en el contexto del Esquema de Inclusión. Dado que el mayor problema a la hora de construir un argumento diagonal o de reducir una paradoja al Esquema de Inclusión es siempre encontrar el operador δ apropiado, un estudio de las diversas maneras que tenemos de construir operadores de este tipo resultará sin duda valioso a la hora de entender el problema de la formación de paradojas de inclusión.

II SOLUCIONES A LAS PARADOJAS DE INCLUSIÓN

"*Two gardeners were working in Nasr al-Din's garden. One of them was tending Nasr al-Din's cabbages and, finding snails, he began to kill them and throw them over the wall of the garden. The second gardener approached and asked 'What are you doing?' to which the first gardener replied 'Killing these snails'. 'Why are you killing them?' asked the second gardener, and the first gardener answered 'Because they are eating Nasr al-Din's cabbages'. But the second gardener said 'Let them be; they're not doing much harm, and after all, they have their needs too'. The gardeners continued arguing and began to fight. Nasr al-Din approached, accompanied by his wife. 'What are you fighting about?' Nasr al-Din asked the gardeners, 'Tell me and I shall give my judgement'. The first gardener said 'I say that these snails should be destroyed because they are eating your cabbages.' And Nasr al-Din replied 'You are right'. But the second gardener said 'I say that the snails should be let be, and allowed to meet their needs'. And Nasr al-Din replied 'You are right'. Then Nasr al-Din's wife said to Nasr al-Din 'But Nasr al-Din, they cannot both be right'. And Nasr al-Din replied 'You are right'.*"
(Cuento tradicional, extraído de Priest 1987, p.155)

4 El Principio de Solución Única, enfoques consistentes y "parametrización"

4.1 Introducción

En anteriores capítulos estudiamos un grupo de paradojas popularmente conocidas como paradojas de autorreferencia o reflexividad. Vimos que todas ellas son en el fondo ejemplos de un tipo de razonamiento genérico al que denominamos contradicción de inclusión y cuya estructura expusimos en forma de esquema. Más allá de este hecho puramente descriptivo, intentamos explorar cuáles eran las virtudes explicativas del Esquema de Inclusión (EI). Defendimos en el capítulo anterior que los elementos principales de dicho esquema, Ω y δ, nos ayudan a comprender mejor el problema de fondo en este tipo de razonamientos. Todos ellos involucran una totalidad absoluta Ω (el conjunto de todos los "φs") y un diagonalizador δ. Aquí, δ es un operador que tiene por límite iterativo a Ω y que, a partir de todos los elementos de Ω, consigue generar circularmente un elemento, $\delta(\Omega)$, que es φ y pertenece a Ω (*clausura*), pero que forzosamente se sitúa fuera de Ω (*trascendencia*), transgrediendo, así, su propio límite iterativo. Vimos cómo intervenía cada elemento de EI en la demostración de diversas contradicciones de inclusión y cómo podíamos clasificarlas apelando al papel desempeñado por ψ en ellas y al tipo de estrategia empleada en las demostraciones de la condición de trascendencia. Apuntamos, así mismo, el estrecho vínculo existente entre las contradicciones de inclusión y ciertos problemas filosóficos recurriendo a la clasificación que Priest hace de éstos en términos de "contradicciones en los límites de la iteración, la concepción, la cognición y la expresión".

Afirmamos, también en el capítulo anterior, que ciertas intuiciones estructurales que tradicionalmente asociamos (de forma acertada) a las paradojas reflexivas, como por ejemplo la presencia de cierto grado de circularidad viciosa y la relación con argumentos diagonales, se ven reflejadas en el Esquema de Inclusión.

Uno de los objetivos centrales de este capítulo (y del siguiente) es estudiar en qué sentido cuestiones estructurales como las que hemos estudiado pueden tener relevancia a la hora de buscar una solución para las contradicciones de inclusión paradójicas. ¿Conlleva el hecho de que todas ellas compartan la misma estructura que hayan de ser solucionadas del mismo modo? Examinaremos aquí varias propuestas teóricas y nos serviremos de EI para identificar estrategias de solución recurrentes. Veremos que la mayoría de soluciones nos enfrenta habitualmente al dilema de elegir entre "consistencia" y "generalidad". Bajo el rótulo de "soluciones paramétricas", estudiaremos un tipo de estrategias que busca bloquear las paradojas construyendo un marco teórico donde, voluntaria o involuntariamente, se sacrifica la "generalidad" (expresiva, deductiva, etc.) en aras de la "consistencia". En el siguiente capítulo, estudiaremos también la solución que Priest da a las paradojas. Veremos hasta qué punto salva los problemas con que tropiezan las soluciones alternativas (discutidas aquí) y qué lectura hace de la relación existente entre la solución a las paradojas reflexivas y su estructura compartida.

4.2 Condiciones para una solución satisfactoria de las paradojas reflexivas: el PSU

Cuando nos enfrentamos a una paradoja, son básicamente dos los intereses que nos guían. En primer lugar, queremos comprender cuál es el problema de fondo, llevar a cabo un "diagnóstico" de la situación que nos permita identificar sus factores esenciales y ver por qué se genera la paradoja. En segundo lugar buscamos, con fines "terapéuticos", encontrar una solución que bloquee la inferencia de algo absurdo.[1] Con respecto a las paradojas reflexivas, parece claro que el Esquema de Inclusión puede ser de gran ayuda a la hora de efectuar un "diagnóstico" o, al menos, eso intentamos mostrar en el capítulo anterior. Allí utilizamos el esquema no sólo para describir similitudes estructurales de carácter general sino también para localizar divergencias estructurales relevantes en la interpretación que hacíamos, en cada caso particular, de los principales elementos de EI. Lo que allí nos importaba sobre todo era ver *cómo* y *por qué* se producen en general las contradicciones de inclusión. Quizá clasificarlas en "subfamilias" disjuntas no resulte factible a raíz de las dificultades con que tropezamos entonces. Sin embargo, la gran virtud de EI es que, al identificar los ingredientes *básicos*, δ y Ω, que intervienen en *todos* los razonamientos estudiados, nos ayuda a ver con más claridad el núcleo de toda contradicción de inclusión: $\delta(\Omega)$. Pudiendo así describir qué peso específico tienen δ y Ω en cada caso y cómo se articulan para

[1] La distinción entre efectuar un "diagnóstico" del problema que suponen las paradojas y prescribir una "terapia" fue introducida en esos términos por Chihara (1979) en el contexto de la paradoja del mentiroso. Un buen número de autores se hacen eco de ella:

producir una contradicción. De esta manera, cabe incluso identificar tipos estándar de demostraciones de contradicciones de inclusión (lo más parecido que tenemos a una taxonomía en términos de "subfamilias"). La pregunta clave ahora es, ¿ofrece este conocimiento algún tipo de ayuda a la hora de solucionar las paradojas, esto es, a la hora de satisfacer nuestros intereses "terapéuticos"?

Si todas las paradojas que nos han ocupado muestran la misma estructura cuando alcanzamos cierto nivel de abstracción y son, por tanto, ejemplos concretos de un mismo problema general de fondo, cabría esperar que podamos decir sobre ellas muchas cosas con validez general apoyándonos en el Esquema de Inclusión. En particular, es tentador pensar que, si el problema "fundamental" es el mismo, una solución a dicho problema aportará, a la vez, una solución para cada una de las paradojas relevantes. Tendríamos así una solución *uniforme* para todas ellas o una estrategia *común* para abordarlas. La generalidad del problema al que nos enfrentamos parece exigir una solución que tenga, al menos, la misma generalidad. Priest expresa esta intuición del siguiente modo:

> "If two paradoxes are of different kinds, it is reasonable to expect them to have different kinds of solution; on the other hand, if two paradoxes are of the same kind, then it is reasonable to expect them to have the same kind of solution. Generalising: it is natural to expect all the paradoxes of a single family to have a single kind of solution. Any solution that can handle only some members of the family is bound to appear not to have got to grips with the fundamental

---

Herzerberger (1982a, p. 479), Barwise y Etchemendy (1987, p. 7) y McGee (1991, p. 2), por ejemplo.

issue. Let us call this the Principle of Uniform Solution (PUS): Same kind of paradox, same kind of solution."[2]

El problema radica aquí (como reconoce Priest poco después) en el significado de la expresión "misma clase". Sin embargo aparcaremos de momento esta cuestión y nos centraremos en otra más amplia, pero relacionada con la presente discusión: ¿Qué requisitos debe cumplir una solución satisfactoria de las paradojas reflexivas?[3]

Toda solución ha de resultar *natural*, debe respetar nuestras intuiciones. En este sentido, un primer requisito que debería cumplir es el de *no ser arbitraria*. Las soluciones deben adecuarse al problema "diagnosticado". Una contradicción se puede resolver de muchos modos pero nuestra solución sólo será satisfactoria si aísla las causas reales del problema que nos ocupa, según un diagnóstico acertado, y se dirige a ellas neutralizando sus consecuencias. Arrancar un brazo no es una manera satisfactoria de acabar con un dolor de muñeca. No buscamos el "efecto colateral" de una medida drástica y desastrosa que muestre más nuestra ignorancia que nuestro conocimiento del tema. Lo que buscamos es comprender el "nudo gordiano" para poder deshacerlo, no cortarlo como hizo Alejandro Magno. Un segundo requisito que debería satisfacer

---

[2] Priest 1994a, p. 32. Reimpreso en 2002 en *Beyond the Limits of Thought*, p. 166. En lo sucesivo me referiré a esta obra como *BLT*. Igualmente, citaré *In Contradiction* (Priest 1987) mediante sus siglas, *IC*.

[3] Consideraciones parecidas a las que ofreceré a continuación se pueden encontrar en diferentes sitios, por supuesto, en Priest (*IC*, capítulos 1-3, y *BLT*, capítulos 9-11). Con respecto a la paradoja del mentiroso, son frecuentes este tipo de observaciones en todas las obras que tratan el tema. Dos buenos compendios son Haack (1978, capítulo 8 - especialmente pp. 160-64 de la traducción al castellano de 1982) y, sobre todo, Kirkham (1992, capítulo 9, especialmente pp. 271-75).

toda solución es el de no ofrecer argumentos *ad hoc*. Esto es, no debería ofrecer argumentos que presupongan la verdad de la propia teoría o que no puedan justificarse con independencia de ésta. Si nuestra teoría tropieza con graves problemas que sólo podemos solucionar asumiendo tesis poco intuitivas, necesitaremos justificar dichas tesis sin apelar a la verdad de la teoría. No podemos decir simplemente: "Sólo si estas tesis son verdaderas, nuestra teoría será correcta. Así pues, estas tesis deben ser verdaderas *porque nuestra teoría es correcta*". Tradicionalmente, se ha considerado también que cualquier "solución" a una paradoja ha de buscar un equilibrio satisfactorio entre dos exigencias fundamentales: En primer lugar, deseamos encontrar una solución *coherente* cuya aceptación no introduzca entre nuestras creencias cosas que nos parezcan absurdas o irracionales, como por ejemplo contradicciones. En segundo lugar, queremos que nuestra solución sea *conservadora* con respecto a lo que tenemos buenas razones para aceptar, es decir, deseamos que no nos fuerce a abandonar o a sacrificar sin motivo cosas que aceptamos por razones de peso. Buscamos, en resumen, una teoría sin "efectos secundarios" no deseados, una teoría que *no genere* creencias absurdas y que *no elimine* creencias razonables que percibimos como irrenunciables.

Habida cuenta de que una paradoja suele ser un razonamiento aparentemente válido cuya conclusión es una contradicción, estas dos exigencias se han materializado habitualmente en dos imposiciones concretas: "consistencia" y "generalidad". Toda solución a las paradojas reflexivas ha de ser *consistente*, ha de introducir las reformas necesarias para impedir que la inferencia de una contradicción (algo racionalmente inaceptable) parezca tener visos de corrección. Además, dicha solución ha de conservar el *alcance* o *generalidad* de las teorías relevantes. En la medida de lo posible, nuestras teorías no deben sufrir recortes expresivos, explicativos o inferenciales. Si tenemos una teoría T a partir de la cual podemos justificar A y B, donde A es una proposición verdadera cuyo conocimiento se apoya en razones sólidas y B, en cambio, una contradicción paradójica, toda reforma T' de T que impida la inferencia de B, ha de preservar la inferencia de A y la capacidad que T tenía para expresar y justificar todo aquello sobre lo que se apoyaba nuestro conocimiento de A.

Uno de los aspectos más relevantes de la "generalidad" que deseamos preservar en nuestras teorías es la "reflexividad" expresiva y explicativa de las mismas. Una teoría que resuelva satisfactoriamente las paradojas ha de ser lo suficientemente general como para *justificarse a sí misma* y expresar sus propios conceptos y soluciones de forma coherente. La intuición o argumento sobre el que se apoya esta afirmación es el siguiente: Si es posible ofrecer una solución *definitiva* para las paradojas, entonces debemos hacerlo mediante una teoría en la que éstas *no se reproduzcan*. Para estar seguros de ello, es necesario que podamos encontrar una teoría general T que dé cuenta de la coherencia de todas las teorías involucradas en la solución de las paradojas y, como T estará involucrada en dicha solución, que dé cuenta de sí misma y de su coherencia. Si esta meta no se alcanzase, tropezaríamos con serios problemas epistémicos y de efabilidad.

Supongamos, por ejemplo, que fijamos unos criterios de corrección C para cualquier teoría que solucione las paradojas y construimos una teoría T que elimina las paradojas en una área determinada A. Si no somos capaces de garantizar *dentro* de la teoría T que los conceptos que ésta introduce pueden ser expresados sin incurrir en paradojas, entonces necesitaremos una teoría T' que exprese T y muestre su coherencia de acuerdo con los criterios C, con lo que la cuestión se reproducirá de nuevo, ahora con T'. Si somos incapaces de encontrar una teoría final que dé cuenta de esta sucesión o jerarquía (posiblemente infinita) de teorías y también de *sí misma*, la expresibilidad y la coherencia del conjunto de teorías quedará por demostrar, lo que arrojará dudas sobre su corrección. Se plantea así un problema doble: Por un lado, tenemos una sucesión infinita de teorías en las que cada teoría T_α justifica la coherencia de las teorías T_β situadas en niveles inferiores ($\alpha > \beta$), pero no existe ninguna teoría T_λ ($\alpha < \lambda$, para toda T_α en la jerarquía) que garantice *por completo* la erradicación de las paradojas y justifique la coherencia de *todas* las teorías de la jerarquía, incluida T_λ. Nos enfrentamos así a un problema clásico de *regreso al infinito en la justificación* (no porque la jerarquía sea infinita, sino porque no contiene ningún nivel finito o *transfinito* que nos permita justificarla en su conjunto). Por otra parte, este problema deriva en otro de *expresividad*. Si (por hipótesis) no existiese una teoría "reflexiva" de este tipo y existiese, en cambio, la jerarquía infinita de teorías que hemos mencionado, entonces resultaría imposible *describir coherentemente* dicha jerarquía. Para hacerlo necesitaríamos una teoría coherente T* que nos hablase de *todas* ellas. Si esta teoría pudiese justificar su propia coherencia, entonces sería la teoría reflexiva general que buscábamos y que hemos descartado por hipótesis. Pero si la teoría no es reflexiva, entonces deberá justificar su coherencia apelando a una teoría superior T*'. Ahora bien, esto indica que T* no es sino una teoría intermedia de la jerarquía postulada, i.e. una teoría que no puede explicar o describir teorías superiores ni, por tanto, la jerarquía en su conjunto.

Junto a las dos exigencias especificadas: *coherencia* y *generalidad* (tres, si subrayamos dentro de la segunda la demanda de *reflexividad*), hemos de añadir una más a la vista de la discusión que apuntamos más arriba. Si el Principio de Solución Uniforme (PSU) tiene sentido y, como de hecho ocurre, todas las paradojas reflexivas comparten (desde cierto nivel abstracto de análisis) una misma estructura, entonces debe existir (desde ese nivel) una única solución compartida que aborde el problema de fondo satisfactoriamente, procediendo del *mismo* modo en cada caso particular. Esta última exigencia resulta controvertida y, por su vaguedad, está sujeta a diferentes interpretaciones, sin embargo su peso se deja sentir en la obra de aquellos autores que han creído firmemente en la existencia de un patrón estructural común a ciertos razonamientos paradójicos. Éste es el caso de Priest, pero también el de Russell. En su artículo de 1908, Russell describía un patrón estructural común a todas las paradojas reflexivas (una relación de circularidad viciosa entre cierta totalidad y una de sus partes). A raíz de este análisis, Russell efectúa un diagnóstico del problema mediante su "Principio de

Circularidad Viciosa" y propone una solución uniforme para todas las paradojas apelando a su teoría de tipos (véase 2.2 y 4.5.1).

En las próximas secciones examinaremos en qué medida las soluciones más populares a las paradojas reflexivas consiguen satisfacer estas exigencias. Pero antes de proseguir es pertinente hacer una observación con respecto al PSU. Son pocos los autores que han dado un trato uniforme a todas las paradojas reflexivas. La mayoría distingue entre paradojas conjuntistas y semánticas, asumiendo tácitamente la división de Ramsey. En el primer grupo encontramos lo que Priest denomina contradicciones en los límites de la iteración (incluyendo la paradoja de Russell [2], cuya descripción en estos términos cuestionamos en 3.2.8). En el segundo grupo la paradoja del mentiroso [1] monopoliza casi toda la atención. Proponer "diferentes tipos" de solución para cada grupo, supone violar el PSU. Por este motivo, examinaré en último lugar esta demanda y me centraré primero en las otras exigencias. Con todo, intentaré identificar en ámbitos dispares, siempre que esto sea posible, estrategias comunes o paralelas.

4.3 Diagonalización y estrategias de solución "paramétricas"

Como hemos apuntado en diversas ocasiones, solucionar una paradoja es, en cierto sentido, disolverla, deshacer una "ilusión", mostrar que lo que nos "parecía" correcto (el razonamiento), o verdadero (las premisas), o falso (la conclusión), no lo es en realidad. Las contradicciones de inclusión sólo constituyen un problema, i.e., una paradoja, cuando no pueden ser ubicadas de forma satisfactoria en una *reductio ad absurdum* de alguno de nuestros supuestos de partida. Solucionar una paradoja consiste, pues, en convertir el argumento que la sustenta en una *reductio*, pero no en una cualquiera, sino en una que llegue a parecernos "natural", incuestionable y pertinente. Cuando tenemos éxito en esta tarea, la paradoja se desvanece y deja de parecernos necesario el absurdo que nos impone, ya que identificamos aquello que hemos de rechazar para bloquear la contradicción y entendemos por qué hemos de rechazarlo sin encontrar ya razones para no hacerlo. Deshacer una paradoja dentro de una área determinada no implica, por supuesto, deshacer todas las paradojas del área, no obstante, en la medida en que vayamos desenmarañando "nudos gordianos", tendremos siempre la sensación de que, más tarde o más temprano, les llegará el turno a otros tal vez más enrevesados. El poema de Parménides que vimos en el primer capítulo esconde seguramente una paradoja relativa al término "ser" usado existencialmente. Sin embargo, las consideraciones que Platón y Aristóteles hicieron con respecto al uso de dicho término por parte de Parménides nos hacen ver que formular esa paradoja tal y como lo hizo Parménides le resta fuerza, ya que tenemos motivos para rechazar razonablemente muchos de sus supuestos y pocos o ninguno para aceptarlos. El teorema de Cantor [I], que vimos en 2.7, es otro buen ejemplo. Quizá haya innumerables paradojas que afecten a nuestra concepción del infinito pero eso no impide que podamos comparar de forma razonable magnitudes infinitas mediante relaciones de correspondencia entre sus miembros y establecer, por ejemplo, que hay tantos números pares como números naturales pero menos números naturales que números reales. Si alguna vez abordar esos problemas suscitó razonamientos paradójicos, ya poco importa. El problema tiene (habitualmente) solución, aunque dicha solución produzca aún sorpresa y ciertos aspectos de la misma planteen, a menudo, nuevos problemas (la cardinalidad de colecciones como On y $\wp$(On), por ejemplo, o del conjunto universal, V, y $\wp$(V), que daba origen a la paradoja de Cantor [10]).

Es importante percatarse de la dimensión histórica que tiene percibir un razonamiento como paradójico. Resumiendo lo aquí dicho, podemos destacar tres hechos: (1) Algunas paradojas se resuelven y, cuando esto ocurre, se convierten en elementos de una *reductio* satisfactoria y ampliamente aceptada de algunas de las creencias que habían desencadenado una conclusión absurda. (2) Toda solución acertada acaba con el tiempo por "deshacer" la paradoja, la "ilusión", transformando así nuestra percepción del mundo y del significado "real" de algunos de nuestros conceptos. (3) La solución de una paradoja suele introducir (o reinterpretar), no obstante, una serie de conceptos, distinciones y argumentos que pueden motivar nuevas paradojas, como ocurre de hecho a menudo.

Junto a estas observaciones históricas cabe también destacar consideraciones de tipo práctico. No todos los ámbitos prácticos en los que encontramos paradojas son del mismo tipo. Ésta es una de las ideas que Ramsey quería subrayar (véase 2.2) al distinguir entre paradojas matemáticas y no matemáticas. En determinados ámbitos formales las consecuencias de las paradojas parecen devastadoras y la urgencia de una solución acuciante. Es, curiosamente, en estos campos donde más condescendientes somos con soluciones drásticas. Soluciones que optan por cortar el "nudo" de las paradojas, antes que deshacerlo (aunque continuemos buscando teorías menos agresivas y más satisfactorias). En otros ámbitos no percibimos las paradojas como una amenaza grave y sí, en cambio, cualquier intento drástico de solución que, para librarnos de ellas, mutile sensiblemente nuestra visión del mundo o la expresividad de nuestro lenguaje.[4]

En el caso de las paradojas reflexivas, abordaremos en esta sección tres cuestiones: ¿Existe alguna estrategia típica para enfrentarnos al problema de las paradojas? Si es así, ¿es exitosa, o presenta algún problema general? Y, por último, ¿son satisfactorias las soluciones propuestas para las paradojas más conocidas? Ofreceremos aquí más un mapa del estado de la cuestión y un esbozo de los problemas principales, que una discusión detallada (aunque haremos excepciones). La razón de ello es evitar reproducir las numerosas y conocidas discusiones que uno puede encontrar desarrolladas de forma clara y eficiente en las referencias bibliográficas que aportaremos.

4.3.1 Estrategias "paramétricas": Restricciones de Ω y δ

En el capítulo 2 defendimos que las paradojas reflexivas tienen la forma de contradicciones de inclusión y en el capítulo 3 (véase 3.2.12) afirmamos que el problema de estas paradojas radica en la posibilidad de

---

[4] Nos gustaría destacar la afinidad de buena parte de las observaciones hechas en esta sección con el espíritu del texto de Quine (1976b) mencionado en el capítulo 1 (nota 15).

inferir una contradicción a partir de δ(Ω). Dado que resolver una paradoja es una cuestión de convertir un razonamiento contradictorio en una *reductio* de alguno de sus supuestos de partida, parece claro que, en el caso de las paradojas de inclusión, tenemos dos opciones: O bien (1) negamos la validez de alguna de las reglas lógicas que nos permiten inferir una contradicción a partir de δ(Ω); o bien (2) negamos δ(Ω). Examinaremos primero la opción (2) por ser la más popular (la tradicional), ya que (1) conlleva introducir cambios en el marco de la lógica clásica.

Optar por (2) requiere negar la existencia o definibilidad de δ o de Ω (o, al menos, la corrección de aplicar δ sobre Ω). Ahora bien, resulta difícil negar δ. Después de todo la operación general que describe esta función parece irreprochable: δ resume un conjunto de operaciones computables que podemos aplicar sobre un grupo de cosas de cierto tipo (un grupo X de "φs") para obtener algo nuevo del mismo tipo ("un φ" que no está en X). En la medida en que el grupo X sobre el que opera δ exista y esté bien definido, δ(X) también lo estará. Así pues, aquellos que adoptan la estrategia (2) dirigen sus sospechas hacia Ω o hacia la posibilidad de definir δ(Ω). Lo más habitual es afirmar que Ω no existe o no puede definirse, de donde se sigue la inexistencia o incorrección de δ(Ω). Ésta es la estrategia más seguida frente a las paradoja reflexivas ("conjuntistas" o "semánticas"). Los dos casos más representativos son la conversión de la paradoja de Russell [2] en una *reductio* que rechaza la existencia del conjunto universal[5] y la conversión de (una u otra versión de) las paradojas del mentiroso [1] y de Grelling [8] en una *reductio* respectivamente de la existencia o definibilidad del conjunto de todas las oraciones (o enunciados, o proposiciones, etc.) verdaderas y del conjunto de todas las propiedades heterológicas (aquellas que no se "autosatisfacen", que no son verdaderas de sí mismas). El más célebre ejemplo de una *reductio* de este tipo es lo que se ha dado en llamar "Teorema de Tarski".[6]

Se trata aquí, no de negar la existencia o definibilidad del conjunto de todas las verdades (o propiedades heterológicas) expresables en un lenguaje cualquiera, sino de negar su existencia o definibilidad *sólo* en lenguajes en los que se puedan formular paradojas semánticas. Un lenguaje que pueda expresar *cualquier* idea susceptible de ser expresada lingüísticamente y en el que pudiésemos traducir los términos y expresiones de cualquier otro lenguaje podría expresar este tipo de paradojas y, por tanto, estaría sujeto al teorema de Tarski. Siguiendo a Tarski, llamaremos "universales" a este tipo de lenguajes sin restricciones expresivas, ideal al que aspirarían, entre otros, los llamados lenguajes naturales o coloquiales (castellano, catalán, inglés, etc.). Si algo

[5] Se puede constatar en cualquier manual de teoría de conjuntos. Por ejemplo, Halmos 1960, capítulo 2 (pp. 15-6 de la traducción al castellano) o Devlin 1993, pp. 35-8. En 1908, tanto Russell (en el artículo citado) como Zermelo (véase *BLT*, p. 158) ya apelan a esta *reductio* para rechazar la existencia del conjunto universal.

[6] En su artículo de 1933, el "Teorema I" (pp. 247-8 de la traducción inglesa) es el que hace más explícita esta *reductio*. Se pueden encontrar exposiciones más sencillas de este tipo de resultados relacionados con Tarski en McGee 1991 (p. 25, teorema 1.3) o, con respecto a las paradojas de Grelling [8] y el mentiroso, en Simmons 1993 (pp. 18; 35-36). También se utilizan resultados relacionados con éstos para construir versiones del teorema de Gödel (véase Smullyan 1992, capítulo 2, pp. 14-27).

se puede expresar, entonces se puede expresar mediante ellos.[7] El "teorema de Tarski" sólo afectaría, pues, a lenguajes capaces de expresar paradojas como la del mentiroso [1] o la de Grelling [8].

Al igual que las paradojas mencionadas, el resto de paradojas que hemos examinado en este trabajo (o, al menos, su inmensa mayoría) se resuelve habitualmente mediante la conversión de su argumento principal en una *reductio* de la existencia o la definibilidad de Ω bajo ciertas condiciones: se niega, por ejemplo, que On y la Jerarquía Acumulativa de Conjuntos sean conjuntos o que, en las paradojas semánticas, Ω exista o esté bien definido.[8] Sin embargo, negar la existencia, definibilidad o descriptibilidad de Ω no es sencillo. A diferencia de lo que ocurre con otras contradicciones de inclusión que no son tenidas habitualmente por paradójicas (y en las que a menudo no necesitamos negar Ω para bloquear $\delta(\Omega)$, sino un supuesto menor, como ocurría en el teorema de Cantor –véase 2.7– o en la "paradoja" del barbero [18] –véase 3.2.9, nota 27–), en el caso de las paradojas reflexivas el sacrificio que nos impone la *reductio* es doloroso.

Renunciar a Ω en el contexto de las paradojas mencionadas (Russell [2], el mentiroso [1] y Grelling [8]) supone renunciar a la

[7] "A characteristic feature of colloquial languages (in contrast to various scientific languages) is its universality. It would not be in harmony with the spirit of this language if in some other language a word occurred which could not be translated into it; it could be claimed that 'if we can speak meaningfully about anything at all, we can also speak about it in colloquial language.'" Tarski, 1933, p. 164 de la traducción inglesa.

[8] En el caso de la paradoja de la cita [11], veíamos en 2.6 que cabía otra solución: rechazar el carácter funcional de las comillas (aunque, entonces, la paradoja se reproduce al considerar en *L* la expresibilidad de *cualquier término funcional* que sirva para formar nombres de expresiones de *L*).

validez *general* de dos principios profundamente arraigados en nuestra concepción ordinaria de "conjunto", "satisfacción" o "verdad". Llamaremos a estos principios "Principio Esquemático de Abstracción" (PEA) y "Principio Esquemático de Satisfacción" (PES) (éste último se encuentra en la base de la definición tarskiana de la verdad y de la Convención o Esquema-V).

(PEA) $x \in \{y: \varphi(y)\} \leftrightarrow \varphi(x/y)$

(PES) x satisface '$\varphi(y)$' $\leftrightarrow \varphi(x/y)$

(φ es aquí una fórmula con una sola variable libre, y, que es sustituida por x en el extremo derecho del bicondicional). PEA dice que x pertenece al conjunto de todos los φs si, y sólo si, x es φ y PES dice que x satisface la propiedad φ si, y sólo si, x es φ.[9] Estos principios no son en sí problemáticos, al contrario, parecen más bien obvios, e incluso *a priori* pero, cuando consideramos el alcance de sus variables de forma no restringida, generan serios problemas. Cuando interpretamos la propiedad φ, PEA, por ejemplo, determina las condiciones necesarias y suficientes para la pertenencia a un conjunto dado. Es bastante natural, pues, asumir que *todas y cada una* de sus ejemplificaciones determinen un conjunto existente (aunque sea el conjunto vacío), sin embargo, el conjunto de Russell, R = $\{y: y \notin y\}$ (una ejemplificación de φ), sumado a PEA, nos

[9] Priest subraya el vínculo obvio que el bicondicional introduce entre ambos principios (*IC*, p. 12). El conjunto que describe PEA es la extensión de la propiedad φ descrita por PES. Este vínculo entre ambos principios es lo que permite describir muchas paradojas conjuntistas en términos semánticos y viceversa (y lo que permite que EI ofrezca una presentación esquemática unificada de todas ellas). Priest defiende la estrecha correlación existente entre las ideas de "satisfacción de una propiedad" y "pertenencia a un conjunto", véase 2.8.3.

impone: $R \in R \leftrightarrow R \notin R$.[10] Del mismo modo, cabe suponer que *cada* ejemplificación de φ en PES ofrezca condiciones necesarias y suficientes para determinar una propiedad (aun en casos en los que nada cumpla dicha propiedad). Sin embargo, una de estas ejemplificaciones: h = '*x* no satisface *x*' ('h' nombra la propiedad 'heterológico'), en conjunción con PES, establece: h no satisface h $\leftrightarrow$ h satisface h. Un problema idéntico se plantea con respecto a otro principio difícilmente cuestionable, el Esquema-V tarskiano:

(PEV) '*x*' es verdadero $\leftrightarrow$ *x*.[11]

(donde *x* es una variable a sustituir por una oración). En teoría, dada una oración *cualquiera*, PEV nos ofrecería las condiciones necesarias y suficientes de su verdad. Sin embargo, existe una oración (la del mentiroso [1]) que dice de sí misma que es falsa: μ = 'μ es falsa' y de la que se sigue, mediante PEV, que μ es equivalente a su negación.

En todos estos casos comprobamos que algunos objetos aparentemente apropiados para interpretar las variables de PEA, PES y

---

[10] Así lo expresa Quine en el caso de PEA (1976b, p. 11): "For any condition you can formulate, there is a class whose members are the things meeting the condition. [...] This principle is not easily given up. The almost invariable way of specifying a class is by stating a necessary and sufficient condition for belonging to it. When we have stated such a condition, we feel that we have 'given' the class and can scarcely make sense of there not being such a class. The class might be empty, yes; but how could there be no such a class at all?" No obstante, concluye: "Yet such exhortations avail us nothing in the face of the antinomy [la paradoja de Russell], which simply proves the principle untenable."

[11] Estas formulaciones descuidadas de PEV y PES están sujetas a la crítica de Tarski con respecto al uso de las comillas como funciones (paradoja [11]). No obstante, ignoraremos esta cuestión porque, como es conocido, ambos esquemas pueden expresarse prescindiendo de comillas (aunque su contenido pierde claridad) y porque lo que discutimos aquí, precisamente, es cómo reaccionar frente a paradojas de inclusión como [11].

PEV (R, h y μ respectivamente), nos fuerzan a aceptar contradicciones. Rechazar estos principios (i.e., *todas* sus ejemplificaciones) mutilaría intolerablemente nuestro esquema conceptual, así pues, la única salida viable consiste en revisar el dominio de las variables que aparecen en dichos esquemas y restringir su alcance para excluir casos "patológicos". Esta decisión nos lleva a replantearnos nuestra concepción intuitiva de lo que significan expresiones tales como "todos los conjuntos", o "todas las propiedades heterológicas", o "todas las verdades" cuyo contenido identificamos crucialmente con el de Ω en las paradojas relevantes ([2], [8] y [1]). Ahora bien, ¿cómo hacemos esto? ¿Acaso "el conjunto de todos los conjuntos" no es un conjunto, o "heterológico" una propiedad, o la "oración del mentiroso" una oración? Nos sentimos fuertemente inclinados a pensar que R, h y μ son el tipo de objetos de los que deberían dar cuenta principios como PEA, PES y PEV, en ese sentido, somos reacios a excluirlos del alcance de sus variables y de la extensión de términos tales como 'conjunto', 'predicado' u 'oración'. Pero, al mismo tiempo, no estamos dispuestos a aceptar las inconsistencias que de ellos se siguen, ésa es la sustancia de la paradoja. ¿Cómo salvar este escollo?

Quienes optan por soluciones de tipo (2) frente a las paradojas y rechazan la existencia o definibilidad de Ω suelen justificar este rechazo apelando a una estrategia general que Priest denomina "parametrización" (*BLT*, pp. 151-55). Veremos a continuación en qué consiste. Nos centraremos en las paradojas [1], [2] y [8] por ser las más estudiadas y constituir los ejemplos más básicos de paradojas formadas a partir de los conceptos de "conjunto", "verdad" o "satisfacción" que, a su vez, son básicos en la mayoría de paradojas aquí examinadas. Los filósofos que estudian [1], [2] y [8] tratan habitualmente de extender las ideas que sus soluciones aportan a la solución de otras paradojas conjuntistas ([7], [9] o [10]) y semánticas ([4], [5] o [6]). Pero, ¿en qué consiste la "parametrización"?[12]

Las estrategias "paramétricas" buscan reconciliar nuestras intuiciones relativas a la inaceptabilidad de una contradicción con nuestra inclinación a considerar R, h y μ como objetos sobre los que podemos aplicar los principios anteriormente mencionados. Para ello buscan resolver la paradoja mediante la introducción de una "distinción" cuyo objetivo es deshacer algún tipo de malentendido, "ambigüedad" o error categorial inadvertido, pero presente en la formulación de la paradoja. Es esta ambigüedad la que provoca la aporía, una vez corregida, vemos, ya sin sorpresa, por qué R, h y μ no pueden aparecer en ninguna ejemplificación paradójica de PEA, PES o PEV.

La idea central es la siguiente: Tomemos el conjunto Ω que aparece en las contradicciones de inclusión y la propiedad φ que lo define (ser verdadero, o heterológico, o un conjunto). Deshacer una paradoja es una cuestión de mostrar que no existe una única propiedad φ que se corresponda con el símbolo 'φ' que aparece en EI, sino varias propiedades diferentes: φ_0, φ_1, φ_2, ... Así pues, el símbolo 'φ' que utilizamos en EI para definir Ω es *ambiguo* y, por consiguiente, también

[12] En *BLT* (pp. 151-55), Priest se centra en soluciones a la paradoja del mentiroso, pero en *IC* (pp. 47-8) sugiere que encontramos estrategias comparables en el caso de las paradojas "conjuntistas".

'Ω' lo es, no puede haber una única colección Ω sino varias, un Ω_n por cada propiedad φ_n. Obviamente, esta ambigüedad se extiende a todos los ámbitos en los que aparece la propiedad φ y, muy especialmente, a los principios que están en juego. PEA, PES y PEV serán principios diferentes según hablemos de φ_n o de φ_m (para todo n $\neq$ m) y, si es necesario, deberemos restringir su alcance para asegurarnos de que sólo se aplican a objetos de los que tiene sentido preguntar si son o no φ_n (para el n relevante). De este modo, las paradojas se resuelven al percatarnos de que un objeto que es φ_n puede muy bien no ser φ_m. Según EI, una contradicción de inclusión tiene la forma de una demostración de $\delta(\Omega) \notin \Omega$ y $\delta(\Omega) \in \Omega$. Una solución paramétrica nos ofrece razones para pensar que, en el fondo, lo que hemos demostrado es $\delta(\Omega_n) \notin \Omega_n$ y $\delta(\Omega_n) \in \Omega_m$. Es decir, hemos demostrado que $\delta(\Omega_n)$ es φ_m pero no φ_n, lo que no es ninguna contradicción, sino una constatación de que ser φ_m no es lo mismo que ser φ_n. Hay oraciones que pueden ser verdaderas$_n$ (verdaderas en un lenguaje n, o en un contexto n, etc.) pero no verdaderas$_m$ (en un lenguaje m, o en un contexto m, etc.) y entidades que pueden pertenecer a colecciones$_n$ pero no a colecciones$_m$. Así pues, indicando con diferentes subíndices o parámetros las propiedades relevantes podemos deshacer la ambigüedad de los términos que originaron la paradoja y ver que hablamos de cosas diferentes en cada caso. Cuando aplicamos el diagonalizador δ sobre Ω_n obtenemos un x que, efectivamente, no pertenece a Ω_n, pero ya no podemos demostrar que x pertenezca a Ω_n, sino a Ω_m, un conjunto diferente. En resumen, δ conserva la propiedad de tascendencia pero pierde la de clausura *cuando se aplica a una totalidad*

Ω_n, con lo que la contradicción desaparece. El diagonalizador δ se comporta ahora tan impecablemente como el diagonalizador "sucesor" (*suc*). Cuando aplicamos *suc* a un número natural, obtenemos su sucesor, un nuevo número natural que no pertenece al conjunto de sus predecesores, sin embargo, cuando aplicamos *suc* sobre ω (la colección de todos los naturales) lo que obtenemos, $\omega + 1$, *no es* un número natural. El diagonalizador δ va saltando de una totalidad Ω a otra cada vez que se aplica sobre una de ellas. Las totalidades Ω pueden formar una "jerarquía" transfinita ordenada, como ocurre de hecho a menudo pero, en principio, esto no es necesario.

En ocasiones, el empleo de estrategias paramétricas no es del todo obvio. En las paradojas conjuntistas, por ejemplo, no se distingue entre diferentes nociones de "conjunto" sino, a lo sumo, entre diferentes nociones de "clase": "clases propias" (R, On, etc.) y "clases" en general (i.e., conjuntos más clases propias). Las primeras abarcan todas aquellas colecciones que no pueden concebirse consistentemente como conjuntos. Pese a todo, esta maniobra es un claro ejemplo de parametrización, simplemente hemos sustituido el término 'conjunto' por el de 'clase' o 'colección' de uno u otro tipo para expresar la ambigüedad inherente a toda paradoja.

Podríamos decir, a modo de resumen, que aquellos que niegan la existencia o definibilidad de Ω en las paradojas de inclusión refuerzan esta solución de tipo (2) con estrategias paramétricas. La idea es la siguiente: $\delta(\Omega)$ implica una contradicción, por tanto Ω no existe o no es definible. Ahora bien, aunque Ω no exista, existe una serie de conjuntos

Ω_0, Ω_1, Ω_2, ..., distintos de Ω que son consistentes al *no cumplir* δ la condición de clausura ($\delta(\Omega_n) \in \Omega_n$) para ningún Ω_n en la serie. Pero, ¿qué ocurre con las soluciones de tipo (1)? ¿Apelan a alguna forma de parametrización?

Recientemente, las soluciones de tipo (1) han cobrado popularidad en el ámbito de las paradojas "semánticas" (no tanto en el de las "conjuntistas"). Las teorías que adoptan soluciones de este tipo suelen cambiar alguna de las reglas o principios lógicos clásicos que permiten la inferencia de una contradicción. Obviamente, al obrar de este modo se abandona la lógica clásica y se propone un marco lógico nuevo con reglas y principios diferentes. Los cambios más frecuentes hacen referencia a la semántica de nuestro lenguaje. Mientras que la lógica clásica es bivalente (i.e., sólo existen dos posibles evaluaciones para una oración: verdadera o falsa), la mayoría de soluciones de tipo (1) a las paradojas aboga por lógicas polivalentes: Dada una oración cualquiera, existen más de dos posibilidades evaluativas (habitualmente tres: verdadera, falsa y "paradójica"). Esto se justifica normalmente apelando a todo tipo de razones: presuposiciones fallidas, defectos semánticos, errores categoriales, ausencia de valor de verdad [*truth-value gaps*], oraciones no significativas, grados de verdad, etcétera. Al adoptar más de dos valores de verdad las lógicas polivalentes deben reinterpretar el significado de las conectivas clásicas ($\neg$, $\wedge$, $\vee$, $\rightarrow$ y $\leftrightarrow$) y de las reglas de inferencia en que aparecen, ofreciendo así un marco lógico diferente.[13] Pero, ¿se apela a

[13] El ejemplo más conocido de solución de tipo (1) es la que ofrece Kripke (1975) a la paradoja del mentiroso.

algún tipo de "parametrización" en estas soluciones? Pese a no resultar obvio, la respuesta es afirmativa, aunque aquí no se "parametriza" la totalidad Ω que aparece en EI, sino las *lógicas* que ofrecen una solución a las paradojas.

La idea es la siguiente: $\delta(\Omega)$ conduce a una contradicción. Como no queremos rechazar la existencia o la definibilidad de δ, de Ω o de $\delta(\Omega)$, debemos rechazar alguna de las reglas o principios (i.e., el marco lógico) que nos condujo a la contradicción de inclusión. Lo que la paradoja muestra, según la solución de tipo (1), es que existe algo: $\delta(\Omega)$ (esto es, h, μ o R en los ejemplos anteriores), de lo que no podemos dar cuenta coherentemente en el marco lógico, l_0, en el que surgió la paradoja. Así pues, debemos buscar un marco lógico nuevo, l_1, en el que podamos razonar sobre objetos como $\delta(\Omega)$ sin que $\delta(\Omega)$ genere una contradicción. En una lógica trivalente, por ejemplo, la oración del mentiroso (i.e., μ) podría recibir una evaluación distinta de "verdadera" o "falsa" (i.e., "carente de valor de verdad"), de este modo la evaluación de μ no sería contradictoria. Si la nueva lógica, l_1, introduce conceptos que dan lugar a nuevas paradojas de inclusión que no pueden resolverse en l_1, nos veremos forzados o bien a rechazar $\delta(\Omega)$ en la nueva paradoja –i.e., a adoptar una solución de tipo (2)–, o bien a rechazar alguna regla de l_1 –i.e., a adoptar de nuevo una solución de tipo (1)– y aceptar un nuevo marco lógico, l_2, que resuelva la paradoja. Si adoptamos reiteradamente estrategias de tipo (1) frente a las paradojas que introducen las nuevas lógicas, obtendremos una serie de lógicas parametrizadas: l_0, l_1, l_2, ..., en las que el salto de una a otra se justifica siempre apelando a $\delta(\Omega)$ en la paradoja de inclusión relevante. Una de las razones por las que las estrategias de tipo (1) son menos populares que las de tipo (2) es que la idea de una jerarquía de lógicas resulta poco convincente. Frecuentemente quienes optan en primera instancia por una estrategia de tipo (1) para resolver una paradoja recurren a una estrategia de tipo (2) para resolver cualquier paradoja nueva que genere el marco lógico propuesto.

4.3.2 El problema de la diagonalización: paradojas reforzadas

Tanto si la estrategia que empleamos es paramétrica con respecto a φ y Ω, como si lo es con respecto a l (la lógica empleada), este tipo de soluciones tropieza siempre con un obstáculo casi insalvable, un auténtico mata gigantes. Para convertir los argumentos diagonales paradójicos en argumentos diagonales "buenos" hemos decidido introducir distinciones, sustituir algo que parecía único (φ, Ω o l) por una pluralidad de propiedades, colecciones o lógicas que distinguimos mediante subíndices o parámetros (φ_0, φ_1, ...; Ω_0, Ω_1, ...; l_0, l_1, ...). De esta manera, el diagonalizador δ actúa como una operación que "expulsa" el término problemático de una determinada colección, Ω_n, o marco lógico, l_n, y lo sitúa en otra colección, Ω_m, o marco lógico, l_m, distintos. El argumento que desemboca en la paradoja se reinterpreta ahora como una *reductio ad absurdum* de la unicidad de Ω o de la validez general de l.[14]

Ahora bien, las paradojas reflexivas nos hablan siempre de totalidades completamente generales o "absolutas": de *todas* las verdades, de *todos* los conjuntos, de *todos* los ordinales, de *todas* las propiedades heterológicas, de *todos* los ordinales definibles, ... Las soluciones paramétricas sustituyen cada una de esas totalidades por una sucesión (jerárquica o no) de colecciones distintas, la cuestión ahora es: ¿Qué ocurre cuando nos hacemos preguntas *generales* sobre esa sucesión de colecciones? ¿Qué ocurre cuando nos preguntamos si algo es verdadero en *cualquier* lenguaje o contexto; o cuando nos preguntamos si algo es verdadero en *cualquier* interpretación semántica (clásica o no); o cuando nos preguntamos si algo pertenece a *cualquier* tipo de clase (propia o no)? En general, estas preguntas pueden responderse sin problemas de coherencia, pero hay excepciones. Responder a estas preguntas presupone un marco general desde el que podamos hablar de cualquier colección Ω_n, o propiedad φ_n, o lógica l_n de la serie relevante. Si podemos hablar, por ejemplo, de *todas* las colecciones Ω_n, entonces tendremos formas de hacer afirmaciones *generales*, de decir que algo es válido para *cualquiera* de ellas. En particular, podremos reunir en una colección Ω^* todas las colecciones relevantes, $\Omega^* = \Omega_0 \cup \Omega_1 \cup \Omega_2 \cup ...$ apelando a una propiedad disyuntiva $\varphi^*(x) = \varphi_0(x) \vee \varphi_1(x) \vee \varphi_2(x) \vee ...$ Una vez construyamos Ω^* y φ^*, es una cuestión de tiempo y maña

[14] Para acceder a una presentación original de los argumentos diagonales, de su clasificación en "buenos" y "malos" y de su relación con la paradoja del mentiroso,

encontrar un diagonalizador apropiado δ que satisfaga las condiciones de *trascendencia* y *clausura* al aplicarse sobre Ω^*, generando una contradicción. Ω^* es una colección bien definida, puesto que cada uno de los Ω_n, lo es y Ω^* es la unión de todos ellos. Cuando aplicamos δ sobre Ω^* obtenemos un φ^* (i.e., un φ_n para algún n en la definición de φ^*) y, por tanto, un elemento de Ω^*, sin embargo, lo que obtenemos no puede pertenecer a Ω^* por ser δ un diagonalizador. Así pues, tenemos, por un lado, la totalidad *absoluta* Ω^* de los φ^*s y, por otro, una función δ que produce φ^*s que no están en Ω^*. La paradoja está servida.

Otro tanto ocurre cuando consideramos el caso de las lógicas parametrizadas, esta vez se trata de construir una paradoja de inclusión adecuada al marco semántico que proporciona la lógica l_n (aquella desde la que, supuestamente, podemos evitar las paradojas). Para cada oración '*p*', la lógica l_n contendrá una serie de posibles evaluaciones semánticas incompatibles: e_0, e_1, e_2, ... Supongamos que e_0 representa en l_n el valor de verdad "*verdadero*". La versión correspondiente de PEV nos dirá que una oración '*p*' es e_0 (i.e., verdadera) si, y sólo si, es el caso que *p* –presumiblemente, PEV asignará el *mismo* valor semántico a "*p*"y a "'*p*' *es* e_0" si ha de acomodar nuestras intuiciones preteóricas sobre la verdad–. Para obtener una paradoja sólo tenemos que considerar el conjunto de todas las oraciones verdaderas: $\Omega = \{x\text{: } x \text{ es } e_0\}$, y hacer que $\delta(\Omega)$ sea la oración μ, donde μ = 'μ no es e_0' o, alternativamente, $\delta(\Omega) = \mu^*$ = 'μ^* es e_1 o e_2 o ...' En ambos casos tenemos una contradicción: (1)

véase Simmons (1993, capítulos 2-4).

Si μ *es* e_0, entonces PEV establece que *no es* e_0 y, (2) si μ *es* e_1 *o* e_2 *o* ..., entonces *no es* e_0 pero, como eso es lo que afirma μ, PEV establece que μ *es* e_0. Un razonamiento similar muestra que también μ^* es equivalente a su negación.

Obsérvese que al construir Ω^* a partir de las colecciones parametrizadas, no anulamos las distinciones paramétricas introducidas anteriormente, unir colecciones no es confundirlas. Los parámetros no desaparecen, simplemente aparece una *nueva* colección, Ω^*, que reúne a las colecciones Ω_n. En el caso de las lógicas, la situación es diferente, dos lógicas distintas son incompatibles entre sí y, como es obvio, no pueden "unirse" sin más. Sin embargo, nuestro argumento no une conjuntos de lógicas parametrizadas, sencillamente construye una nueva paradoja en el marco semántico de la lógica que "resolvía" la anterior paradoja. Alguien podría rechazar, en el caso de Ω^*, la legitimidad de unir las colecciones parametrizadas en una supercolección que las comprenda a todas pero, para que su queja resultase aceptable, debería ofrecer razones que no fuesen *ad hoc* (i.e., razones independientes de la aparición de la nueva paradoja, puesto que lo que está en cuestión es, precisamente, cómo reaccionar frente a una paradoja). En caso contrario, nos resultaría difícil objetar a la validez intuitiva de una operación que se limita a reunir colecciones con vistas a hacer afirmaciones generales sobre ellas.

Otra posible salida es afirmar que, en contra de lo que creemos, Ω^* no es completamente general, todavía encierra cierta ambigüedad paradójica que debemos desentrañar: hay muchas Ω^*_n diferentes. También esta maniobra debería justificarse sin argumentos *ad hoc*, cosa difícilmente concebible en este caso. Pero, sea como fuere, tampoco esta opción nos llevaría muy lejos porque, lo que nos ofrece nuestro argumento, es una estrategia general para atacar este tipo de reacciones. Sólo tenemos que generar una nueva reunión de conjuntos parametrizados de orden superior y construir un nuevo argumento diagonal del mismo orden. Otro tanto podemos decir con respecto a la parametrización de lógicas, siempre podemos "saltar" a otro marco semántico-lógico para resolver una paradoja, pero nuevas versiones reforzadas de la misma, siguiendo la estrategia esbozada, atacarán la solución propuesta.

Cuando Cantor descubrió los argumentos diagonales al demostrar que había más números reales que naturales, observó que el problema no se resolvía añadiendo a la lista de números reales y naturales emparejados el contraejemplo generado por la diagonalización. Añadir dicho contraejemplo meramente diseñaba un marco en el que podíamos construir un nuevo contraejemplo por diagonalización. Cantor se felicitó por la generalidad de su argumento que, además, podía aplicarse a una infinidad de casos para demostrar que el conjunto potencia de cualquier conjunto A es siempre mayor que A (el teorema [I] de 2.7). La generalidad de los argumentos diagonales (incluso de los que conducen a paradojas) es uno de sus rasgos más característicos. Podríamos insistir incansablemente en nuestra estrategia de encontrar soluciones paramétricas de orden superior (lo que propiciaría incansables contraejemplos por diagonalización con las técnicas descritas), pero si optamos por esta huida hacia delante, lo que peligra ahora es nuestra capacidad para referir coherentemente nuestra solución. Sería difícil hablar en términos *generales* (como observamos en 4.2) de la jerarquía infinita de colecciones paramétricas que se multiplican indefinidamente. Ninguna expresión de nuestro lenguaje sería lo suficientemente general como para hablar de todas ellas, ningún lenguaje podría describir o referir consistentemente toda la jerarquía, tales intentos no serían sino "episodios intermedios" dentro de la misma.

El problema general que aquí hemos descrito es el que plantean a todas las soluciones paramétricas las "paradojas reforzadas", como popularmente se las describe. Básicamente, la objeción afirma que las soluciones paramétricas no erradican por completo las paradojas reflexivas, sólo las aplazan, reubicándolas en una instancia superior dentro del nuevo marco conceptual que utiliza la teoría para resolver las inconsistencias que nos preocupan. A continuación veremos cómo encajan diferentes soluciones a las paradojas reflexivas más famosas ([1] y [2], sobre todo) en este enfrentamiento dialéctico entre "parametrización" y "reagrupación" en el que los argumentos diagonales tienen un peso distinto en cada bando.

4.4 Soluciones a las paradojas "conjuntistas"

Uno de los rasgos característicos de estas paradojas es el alto nivel de consenso y aprobación que existe con respecto a cómo solucionarlas, actitud que contrasta con la falta de acuerdo que reina en el ámbito de las paradojas semánticas. Los sistemas clásicos de teoría de conjuntos (y también los sistemas de conjuntos no fundamentados) han evitado las paradojas tomando una serie de decisiones ampliamente conocidas y aceptadas por la gran mayoría de matemáticos. En esta sección, nos limitaremos a ver brevemente en qué sentido estas soluciones se pueden tildar de "paramétricas" y qué motivos tenemos, según Priest, para sentirnos insatisfechos con ellas.[15]

4.4.1 ZF: Zermelo y la Jerarquía Acumulativa de Conjuntos

La solución de Zermelo a la paradoja de Russell [2] consiste en negar la existencia del conjunto R de todos los conjuntos (Ω en la contradicción de inclusión) y en restringir los axiomas de la teoría de conjuntos que puedan propiciar la inferencia de R o de otros conjuntos problemáticos (como, por ejemplo, On). Las reformas más notables son el Axioma de Fundamentación (para evitar la circularidad) y, sobre todo, la restricción de PEA, que es sustituido por el *Esquema Axiomático de Separación* (EAS). Según EAS no toda condición o propiedad define un conjunto, como presuponía PEA. Sólo definen un conjunto aquellas condiciones o propiedades φ que seleccionan, a partir de un conjunto *previamente dado* A, un subconjunto $X \subseteq A$ ($x \in X \leftrightarrow x \in \{y \in A: \varphi(y)\}$), garantizamos así la existencia de X tomando de A sus elementos. El Axioma de Fundamentación impone, por su parte, que todo conjunto (no vacío) sea disjunto con respecto a alguno de sus elementos, esta

[15] Todas las objeciones que resumimos en este apartado se pueden encontrar en *IC*, capítulo 2, y en *BLT*, capítulo 11. Remitiremos al lector a estas obras en todo momento. Omitiremos las soluciones de Russell y Ramsey a las paradojas "conjuntistas" (la conocida teoría de tipos). Una exposición crítica de las mismas se puede encontrar en *BLT*, pp. 136-40; 148. Sólo destacaremos que la teoría de tipos es el ejemplo que utiliza Priest como paradigma de parametrización.

condición asegura en la práctica que ningún conjunto A pueda pertenecerse a sí mismo ni a ninguno de los conjuntos que aparecen en las cadenas descendentes que, partiendo de A, podemos formar mediante la relación de pertenencia. También los regresos al infinito formados mediante '$\in$' quedan excluidos por el resto de axiomas del sistema ZF (Zermelo-Fraenkel). Estos axiomas permiten describir la totalidad de los conjuntos de ZF como el resultado de aplicar repetidas veces, con el conjunto vacío como punto de partida y a través de niveles transfinitos, las operaciones de "conjunto potencia" ($\wp$) y "unión" ($\cup$) sobre los miembros de niveles inferiores. El resultado es la *Jerarquía Acumulativa de Conjuntos* (JAC) descrita en la paradoja de Mirimanoff [9] (el razonamiento usado en ZF para mostrar que JAC no puede ser un conjunto, i.e., JAC $\notin$ JAC). "Ser un conjunto" significa, pues (en ZF), aparecer en JAC.

Para hacer honor a la verdad, la solución Zermelo-Fraenkel a las paradojas no es "paramétrica", quizá por ello tampoco se preocupa demasiado por reconciliarse con nuestras intuiciones preteóricas sobre la noción de conjunto: se limita a "extirpar" sin más cierto número de entidades. La mayoría de críticas que ha recibido ZF señalan la gratuidad de alguno de los sacrificios que nos impone. En este sentido, el Axioma de Fundamentación, por ejemplo, excluye una serie de conjuntos, los no fundamentados, cuya coherencia y existencia como entidades matemáticas no suponen ningún problema a juicio de muchos.[16] También

[16] Véase 2.3, nota 12, donde hablábamos de pruebas de *consistencia relativa*. En estas pruebas se construye en ZF un "modelo" para los sistemas no fundamentados (como ZFA) que asigna a cada hiperconjunto un conjunto isomorfo en ZF. La existencia de este

Priest rechaza (*IC*, pp. 39-41) algunos argumentos en contra de los conjuntos no fundamentados y de los regresos que estos entrañan. Se opone sobre todo a la idea de que los elementos de un conjunto dado hayan de "preexistir" o preceder a éste, ya sea "temporal" o conceptualmente.

Pero, además de los conjuntos no fundamentados, tenemos razones de peso, según Priest, para aceptar también colecciones como R, On o JAC. Al excluir estas colecciones ZF aporta argumentos *ad hoc* (*IC*, pp. 38-9) y restringe de forma insatisfactoria el universo de los conjuntos. La jerarquía debería motivarse y justificarse con independencia de su éxito a la hora de resolver paradojas (*IC*, pp. 39-41), sobre todo si tenemos buenas razones para aceptar las colecciones que excluye (*BLT*, p. 158). La exclusión de conjuntos no fundamentados es, de hecho, una razón independiente pero, como hemos visto, controvertida, dada la amplia aceptación que dichos conjuntos tienen. Una noción más amplia de conjunto parece conveniente. Priest incluso sugiere que la idea de una "jerarquía acumulativa" presupone una noción de conjunto *diferente* de la que define. Sin una idea previa de ordinal, por ejemplo, es imposible describir la formación de la jerarquía.[17] Si en ZF un conjunto es, por definición, lo que aparece en JAC, la idea intuitiva (y previa a JAC) de conjunto ordinal en que se apoya ZF para construir JAC queda excluida

"modelo" matiza o incluso devalúa cualquier posible crítica desde ZFA (u otros sistemas no fundamentados) a la capacidad expresiva de ZF. Desde "cierta" perspectiva estructural, ambas teorías hablan de las "mismas entidades".

[17] De hecho, algunos manuales, como Devlin 1993, introducen de forma bastante prematura la noción de ordinal para poder explicar intuitivamente, pero con precisión, qué es JAC.

del universo de los conjuntos (*IC*, p. 38-9). Priest insiste también en que nuestras intuiciones sobre PEA no se limitan *sólo* a sus ejemplificaciones *dentro de* JAC. Afectan por igual a conjuntos fuera de la jerarquía (*IC*, p. 39). Ciertamente, "[i]t is a remarkable fact that virtually every consistent theory of sets that has been proposed this century can be shown to hold in an initial segment of the hierarchy" (*IC*, p. 38). Sin embargo, hay excepciones (Priest cita el sistema NF de Quine) que se han visto relegadas a la categoría de meras "curiosidades".

Priest aporta en *BLT* (pp. 158-9) otros apoyos directos en favor de nuestra concepción "naif" de conjunto. Muchas de estas intuiciones se basan en la generalidad de otro principio, el *Principio del Dominio* [*Domain Principle*] (PD) inspirado en ideas de Cantor y citado por otros autores (*BLT*, pp. 123-6). El principio estipula que "*a todo infinito en potencia le corresponde un infinito en acto existente*". PD justifica la existencia de los conjuntos infinitos y, en particular, de los ordinales límite ya que, dada una progresión infinita de ordinales, PD establece la existencia de un ordinal que los contiene a todos.[18] Ahora bien, para que PD no sea una tautología vacía hemos de entender de forma diferente un *infinito en acto* (un conjunto definido, una constante) y un *infinito en potencia* (una cantidad variable e indefinidamente extensible que va más allá de cualquier límite).[19] Para Priest, PD viene a decir que, dada

[18] En ZF la existencia de ω, el primer ordinal y conjunto infinito, se *postula* mediante el Axioma de Infinitud, que asume PD. La existencia del resto de ordinales límite se sigue de la de ω usando, según el caso, varios axiomas: el Axioma de Reemplazo, el Axioma de Conjunto Potencia, etc.

[19] "... a potential infinity is some kind of variable quantity, whose variation can go beyond any preassigned bound of a certain kind." *BLT*, p. 124.

cualquier cantidad variable de *x*s, existe la *totalidad* o el *conjunto de todos los x*s.[20] Las teorías clásicas de conjuntos, siguiendo a Cantor, aceptan la validez de este principio cuando admiten la existencia de *conjuntos infinitos* como ω o $\aleph_7$, sin embargo, rechazan su validez cuando hablan de R, JAC u On. ZF afirma que no existe ningún conjunto (i.e., infinito en acto) que se corresponda con el infinito en potencia que representan R, JAC u On. Priest, en cambio, rechaza esta afirmación por considerarla *ad hoc* y atribuye a Cantor el principal argumento que ofrece a favor de la validez general, no restringida, de PD (*BLT*, pp. 125-6). Según este argumento, entender una oración que habla de una cantidad (finita o infinita) de entidades es entender una expresión lingüística que contiene una variable cuantificada con un *dominio* de interpretación determinado (de ahí el nombre de PD). Si el dominio de interpretación no estuviese determinado en algún sentido (i.e., si no *existiese* el *conjunto* de *todo* aquello que puede ser sustituido por la variable), el sentido de la expresión permanecería indeterminado.

Cuando hablamos de JAC, o de conjuntos, o de ordinales *en general*, utilizamos frases del tipo: 'Para todo *x* –o, Existe un *x*– tal que *x* es un conjunto (o un ordinal, o un elemento de JAC, etc.) ...' Si el dominio de *x* no está determinado, entonces tampoco el significado de las

[20] "I take [PD] to be a formulation of the Kantian insight that totalisation is conceptually unavoidable." *BLT*, p. 124. Priest subraya el paralelismo que varios autores encuentran entre las paradojas del infinito absoluto en Cantor y las antinomias kantianas (*BLT*, pp. 120-1). El punto de encuentro entre ambas es lo inevitable que resulta para nuestra razón considerar la existencia de ciertas totalidades. Priest se esfuerza (tras criticar los argumentos kantianos a favor de la tesis y la antítesis en las cuatro antinomias) por construir una "quinta antinomia" [19] de espíritu kantiano que tiene de hecho la misma

expresiones en que aparece x lo estará. Ahora bien, tanto en nuestra exposición de ZF como de cualquier otra teoría de conjuntos, empleamos continuamente oraciones que tienen por dominio el universo de todos los conjuntos, On o JAC. Si estas totalidades no existiesen de hecho, las oraciones carecerían de sentido.[21] Sin la validez general de PD, tropezamos con serios problemas a la hora de expresar o describir ZF.[22] Priest señala (*IC*, pp. 45-6) algunos problemas técnicos derivados de esta situación en la interpretación de las expresiones lógicas que cabría utilizar en una formulación completa de los presupuestos de ZF. Para empezar, no podríamos describir ZF dentro del lenguaje de ZF, ya que ninguna de sus variables tomaría a R, On o JAC como dominio de interpretación. Así pues, necesitaríamos un metalenguaje en el que poder hablar de ZF (*IC*, pp. 47-8). Pero, ¿qué estatus tendría tal metalenguaje, cómo evitar que se reproduzcan en él las paradojas? Llegados a este punto, la justificación de ZF debe apelar a algún tipo de estrategia paramétrica (como las discutidas en 4.3.1), aún por especificar.

Para finalizar, mencionaré que Priest ofrece otros argumentos más técnicos en contra de ZF. Aduce (*IC*, pp. 41-5), por ejemplo, que ZF no es una teoría satisfactoria desde el punto de vista de la Teoría de Categorías (donde se busca describir estructuras matemáticas generales y

---

estructura que las paradojas del infinito absoluto: una contradicción de inclusión en los límites de la iteración.

[21] Obviamente, existen muchas maneras distintas de entender la cuantificación de variables. Priest considera posibles objeciones a su argumento en *BLT* (pp. 126, 159-63 –se centra sobre todo en modelos intuicionistas de cuantificación–). No reproduciré aquí los argumentos en una y otra dirección.

[22] Varios comentaristas critican la aceptación de PD, Priest responde a sus críticas en *BLT*, pp. 280-2.

descubrir isomorfismos entre entidades aparentemente diferentes), ZF carece de generalidad suficiente para expresar sus resultados (aunque se consideran objeciones posibles a esta afirmación). Priest también critica el hecho de que en ZF las paradojas conjuntistas aparezcan completamente "divorciadas" de las semánticas y, en particular, que no se haga hincapié en los problemas de coherencia y expresividad que plantea describir aquello que, sin existir en ZF, constituye su objeto de discurso: la totalidad de los conjuntos (JAC, en este caso). Las obras de Gödel y Tarski han puesto de manifiesto la estrecha relación que existe entre ciertas "paradojas" matemáticas (Priest interpreta los teoremas de incompletud de Gödel en este sentido –*IC*, capítulo 3, pp. 49-63) y nociones semánticas formalizables.

4.4.2 Von Neumann y las clases propias

Existe un modo de conservar, por un lado, la idea de conjunto que reivindica ZF y de respetar, por otro, nuestras intuiciones con respecto a la generalidad de PEA, PD y la existencia de colecciones como R, On o JAC (que ZF presupone): A saber, recurrir a algún tipo de solución paramétrica que introduzca distinciones con respecto a los objetos que aparecen en ejemplificaciones de PEA y PD. La solución más conocida en este sentido es la distinción de von Neumann entre "Objetos-I" y "Objetos-II", según hablemos respectivamente de colecciones que pueden ocurrir en la parte izquierda o en la parte derecha de enunciados de la forma "$x \in y$".[23] Estas categorías no son mutuamente excluyentes, algunas colecciones pertenecen a ambas, otras sólo a una. Los objetos-I son, básicamente, los conjuntos de ZF, los objetos-II reciben el nombre de "clases". Todos los conjuntos (salvo ∅ y, en caso de que los haya, los urelementos) son además clases y muchas clases son conjuntos, aunque no todas. Las clases que no pueden ser conjuntos reciben el nombre de "clases propias". Por definición, una clase propia *no puede pertenecer* a ninguna entidad. Von Neumann identifica como clases propias a todas aquellas clases que sean al menos *tan grandes* como JAC (i.e., On, R, y el resto de colecciones paradójicas). Las paradojas se resuelven haciendo que sea imposible considerar si clases propias como R, On o JAC pertenecen o no a objetos de cualquier tipo (ni siquiera a ellas mismas o a otras clases propias). Esta solución no renuncia a la existencia de las colecciones paradójicas, ni tampoco al discurso que versa sobre ellas, simplemente las distingue de los conjuntos.

Pero tampoco esta solución satisface a Priest, que apunta de nuevo su carácter *ad hoc*. No se nos ha ofrecido ninguna razón, con independencia de las paradojas que provocan, para establecer que las clases propias no puedan pertenecer a otros objetos. Von Neumann sugiere que estas clases son "demasiado grandes" para formar parte de otras clases pero no esta claro que el problema que aquí nos ocupa sea un

[23] Presento la idea tal y como lo hace Priest en *BLT*, existen ciertas divergencias con respecto a la formulación original de Von Neumann, véase *BLT*, p. 163. La distinción de Cantor entre multiplicidades consistentes e inconsistentes anticipa en cierto modo la de von Neumann (*BLT*, p. 164).

problema de "tamaño".[24] En particular, es difícil ver "why size, as such, should have anything to do with the possibility of being a unity, i.e., the possibility of being a member [...] Indeed, to refer to a collection, even if it is a proper class, is, in a certain sense, to treat it as an entity, a unity; *a fortiori*, if one quantifies over it. Any collection (set or class) must be a candidate for membership" (*BLT*, p. 164).

Pero hay una objeción más grave (*BLT*, p. 165). La solución de von Neumann padece los mismos problemas que cualquier solución paramétrica. Dado el nuevo marco conceptual, podemos formular versiones reforzadas de las paradojas. En vez de considerar el conjunto de todos los conjuntos, o de todos los ordinales que son conjuntos, etc., podemos considerar ahora la clase de todas las clases, o de todas las clases que son ordinales. Las versiones correspondientes de las paradojas de Russell [2], Burali-Forti [7], Mirimanoff [9], etc., serían interpretadas una vez más como una *reductio ad absurdum* de la *existencia* de las clases problemáticas. Así pues, poco ganamos con aceptar la existencia de R, la *clase* de todos los conjuntos, si nos vemos forzados a rechazar a continuación R', la clase de todas las clases. La paradoja no ha desaparecido, se ha reubicado y, para evitarla, hemos acabado violando PD al rechazar la existencia de R'. Podríamos solucionar el problema apelando a un nuevo marco de orden superior (i.e., introduciendo nuevas distinciones entre clases de orden 1, 2, 3, ...), pero esto nos remite,

[24] El recurso al "tamaño" para distinguir unas clases (las paradójicas) de otras no parece estar bien motivado o ser plenamente satisfactorio. Varios autores discuten este tema, véase por ejemplo Hallet 1984, o Barwise y Moss 1996, pp. 319-21 (quienes, al igual que Priest en *BLT*, citan a Hallet 1984).

obviamente, a los comentarios que hicimos en 4.3.2, siempre tendremos nuevas versiones de la paradoja y problemas expresivos o de consistencia para referirnos en general a la jerarquía de clases de diferentes tipos. Por último, acaba puntualizando Priest, tampoco esta solución ofrece un marco satisfactorio para la Teoría de Categorías por las mismas razones que ZF.

4.4.3 Conclusión

Tomados de uno en uno, quizá ninguno de los argumentos que ofrece Priest contra ZF (con o sin clases) sea concluyente. Siempre es posible cuestionar la generalidad de PEA, PD y la existencia de R, On o JAC. Sin embargo, creo que tomadas en su conjunto, las objeciones de Priest consiguen al menos tres objetivos: En primer lugar, muestran que, con independencia de las paradojas, no tenemos muchos argumentos sólidos para aceptar las restricciones que introduce ZF. En segundo lugar, muestran que tenemos poderosas intuiciones a favor de la generalidad de PEA, PD y de la existencia de los conjuntos problemáticos, rechazarlos no es en absoluto sencillo o satisfactorio. Y, en tercer lugar, muestran que la propia comunidad científica no está tan contenta como parece con las soluciones que plantea ZF. Existen problemas de expresividad que se ponen de manifiesto en la voluntad de mejorar ZF o de encontrar alternativas más satisfactorias a las restricciones mencionadas (el recurso a la noción de clase es un ejemplo, las necesidades que introduce la teoría de categorías, otro). Por todo ello, cabe afirmar que las paradojas conjuntistas no son, ni mucho menos, fantasmas del pasado.

Sencillamente somos más estrictos a la hora de atajarlas por el carácter formal de la disciplina en que se producen.

4.5 Soluciones a las paradojas "semánticas" (la paradoja del mentiroso)[25]

En el caso de paradojas semánticas como la del mentiroso [1], el consenso con respecto a su solución es prácticamente nulo. Todas las propuestas tienen virtudes innegables y todas ellas tropiezan con problemas –en el fondo muy similares– que sus rivales se encargan de destacar. Ofreceremos a continuación un breve resumen de las teorías más conocidas y respetadas y de las dificultades a las que se enfrentan.[26]

4.5.1 Russell y Tarski: Las jerarquías de tipos y lenguajes

La primera solución que mencionaremos es la que proporciona Russell en 1908 al formular su teoría de tipos, que tiene por objetivo preservar el Principio de Circularidad Viciosa (PCV) –el diagnóstico de Russell al problema de las paradojas reflexivas, véase 2.2. Esta teoría

[25] El problema de la paradoja del mentiroso es independiente del de los portadores de la verdad, se plantea tanto para oraciones como para proposiciones, enunciados, etcétera. Salvo en aquellos casos en los que realmente importe (4.5.3, por ejemplo), seremos ambiguo con respecto a si PEV se refiere a oraciones, a proposiciones o a otras cosas (aunque la mayoría de veces hablaremos de oraciones para simplificar, siguiendo la práctica dominante).

[26] Nos limitaremos a esbozar y situar diversas teorías y críticas en el marco de las observaciones hechas en 4.3.1 y 4.3.2. Ubicaremos bibliográficamente tanto unas como otras. En Kirkham 1992 (pp. 271-306) y Visser 1989 encontrará el lector un resumen de muchas de ellas. Un excelente compendio de críticas (basadas en argumentos diagonales) a las soluciones existentes se puede encontrar en Simmons 1993 (pp. 45-82). En Sainsbury 1995 (pp. 107-33), Gupta y Belnap 1993 (pp. 1-32), *IC* (pp. 11-33) y *BLT* (pp. 136-40, 148-55) también se pueden encontrar interesantes análisis y discusiones críticas de diversas teorías.

constituye un ejemplo paradigmático de parametrización. Apoyándose en PCV, Russell rechaza la existencia de las totalidades Ω en las contradicciones de inclusión y construye una jerarquía de objetos de distinto orden. En esta jerarquía, Russell toma como objetos funciones y proposiciones (reduciendo la idea de conjunto, por ejemplo, a la de función proposicional) y añade a cada variable del lenguaje un índice que determina el orden de los objetos por los que puede ser sustituida. El orden de un objeto es el inmediatamente superior al de la variable de orden más alto que contenga (cero si no contiene ninguna). Las paradojas se evitan prohibiendo que un objeto de orden n verse sobre objetos (i.e., contenga variables u objetos) de orden n o superior, no existen por tanto proposiciones que versen sobre la *totalidad* de las proposiciones, de este modo aseguramos la validez de PCV. (Para más detalles y referencias, véase por ejemplo *BLT*, pp. 136-40; Sainsbury, 1995, pp. 124-29; y, por supuesto, Russell 1908). Una de las virtudes de esta propuesta es que respeta PSU aportando una solución común a todo tipo de paradojas reflexivas, conjuntistas o semánticas. El mayor inconveniente que afronta son las serias restricciones expresivas que impone en ambos campos, matemáticas y semántica. Priest apunta (*BLT*, pp. 137-40) que la solución de Russell sacrifica muchos resultados válidos en teoría de conjuntos (aun después de complementarse con el "axioma de reducibilidad") y muestra ciertos problemas de coherencia interna. La mayoría de críticas que esta teoría recibe hacen referencia a sus limitaciones expresivas: Casos de autorreferencia o referencia cruzada inocuos son desterrados; incluso la propia formulación de la teoría peligra por el argumento esbozado en 4.3.2. Si quisiéramos ofrecer una descripción completamente *general* de la teoría de tipos y expresar sus axiomas (como hace Russell), no podríamos determinar el orden de las variables que aparecen en dichos axiomas. Si esas variables tuviesen algún orden, estaría por encima del orden de las variables de *cualquier* otra proposición (¡incluidos los axiomas!) y, contradiciendo la teoría de tipos, tendríamos variables cuyo dominio sería el conjunto de *todas* las proposiciones. Esto facilitaría la formación de paradojas reforzadas.[27] Así pues, o la teoría es consistente pero inexpresable o, expresable pero paradójica.

Quizá la solución a las paradojas de Grelling [8] y el mentiroso [1] más celebrada e influyente haya sido la que ofreció Tarski (1933) al proponer su teoría de la verdad. Durante décadas sus ideas se presentaron como la salida adecuada al problema. El objetivo de Tarski era construir una teoría que, por un lado, fuese *consistente* (i.e., estuviese libre de paradojas) y, por otro, *materialmente adecuada* (i.e., garantizase la satisfacción de todas las ejemplificaciones de PEV). Como vimos en 4.3.1, la oración del mentiroso, μ, muestra que este objetivo no se puede cumplir en lenguajes que la contengan. Para Tarski (1933, pp. 164-5; 1944, pp. 284-6), el origen del problema se encuentra en los lenguajes "semánticamente cerrados", i.e., aquellos que, como los coloquiales, contienen su propio predicado de verdad y medios para nombrar a sus propias expresiones. En estos lenguajes siempre es posible construir una oración como μ y demostrar una contradicción a partir de PEV y las leyes

[27] Russell intenta evadir críticas similares hablando de "ambigüedad sistemática" en la interpretación de las variables, Priest muestra por qué este parche no funciona (*BLT*, pp.

básicas de la lógica clásica. Ése es su diagnóstico. Consecuentemente, Tarski (1933, p. 165) renuncia a resolver el problema en el caso de los lenguajes naturales, pero proporciona una teoría consistente y materialmente adecuada para lenguajes formalizados no semánticamente cerrados, alegando que el contenido de la mayoría de las ciencias se pueden expresar en este tipo de lenguajes. Su solución le lleva a construir una jerarquía de lenguajes interpretados con distintos predicados de verdad. Ningún lenguaje contiene su propio predicado de verdad (el primer lenguaje ni siquiera contiene el predicado 'verdadero'). Cada lenguaje, L_n, de la jerarquía es "semánticamente más rico" que sus predecesores, L_m (m < n), ya que L_m (o una traducción) está incluido en L_n, que además contiene (a diferencia de L_m) medios para referirse a las expresiones de L_m y un predicado 'verdadero-en-L_m' para sus oraciones. De este modo, cada lenguaje puede considerarse un "metalenguaje" con respecto a los lenguajes que lo preceden y que toma por objeto. La paradoja se resuelve aquí porque no hay ninguna oración en lenguaje alguno de la jerarquía que pueda afirmar de sí misma que es falsa (o no verdadera): la oración μ no es expresable.

Una vez más, la solución "jerárquica" de Tarski es atacada por sus limitaciones expresivas. Un gran número de oraciones autorreferentes inofensivas son excluidas, junto con μ, de los lenguajes jerarquizados. También aquí tropezamos con problemas a la hora de describir la jerarquía, ya que no existe ningún lenguaje desde el que podamos efectuar una descripción completa. Si existiese tal lenguaje, podríamos

138-9).

describir en él *todos* los lenguajes de la jerarquía y construir un predicado de verdad para *todos* ellos. Ahora bien, este lenguaje debe pertenecer a la jerarquía, lo que plantea un dilema: o es semánticamente cerrado y, por tanto, inconsistente; o no es semánticamente cerrado y precisa un metalenguaje distinto y más rico para expresar su predicado de verdad, contradiciendo la hipótesis de partida. Ésta es una nueva versión de la objeción estándar de 4.3.2 (se puede encontrar formulada, por ejemplo, en Soames 1999, p. 155).

Una de las objeciones más demoledoras a cualquier intento de resolver las paradojas semánticas en los lenguajes naturales (e incluso en los formales) apelando a la existencia en ellos de una jerarquía implícita de lenguajes y términos de distinto orden proviene de Kripke (1975). Kripke muestra que estas soluciones ponen en peligro, no sólo la expresibilidad de la propia jerarquía, sino la de un sinfín de oraciones triviales con las que cuenta cualquier lenguaje mínimamente expresivo. La razón de ello es que resulta imposible determinar *sólo* sobre la base de criterios *sintácticos* o *semánticos formalizables* cuándo una oración es paradójica y cuándo no. Los factores empíricos juegan un papel decisivo en este sentido. Según su contexto de uso, una oración fácilmente evaluable en condiciones normales puede convertirse en paradójica (como la aseveración de Epiménides, el cretense: "Los cretenses mienten siempre"). Este hecho muestra que el número de oraciones que deberíamos sacrificar en nuestro lenguaje para introducir una jerarquía "a la Tarski" (i.e., basada en criterios sintácticos) es inmenso, dada la ingente cantidad de oraciones susceptibles de expresar una paradoja en algún contexto. Pero tenemos aquí un problema adicional, consideremos las siguientes oraciones:

(D) Nada de lo que dice Nixon sobre el Watergate es verdadero.

(N) Nada de lo que dice Dean sobre el Watergate es verdadero.

Supongamos que, en el mismo contexto, Nixon profiere (N) y Dean (D). Aun cuando no existan razones para considerar que (N) o (D) sean paradójicas (aun cuando una o ambas sean falsas), existe un problema con la jerarquía. En *este* contexto, el cuantificador en (N) afecta a *todas* las afirmaciones de Dean sobre el Watergate y el de (D) a *todas* las de Nixon sobre el mismo tema. En la jerarquía tarskiana, (D) debería situarse en un lenguaje superior jerárquicamente al de todos los lenguajes en los que Nixon ha hablado sobre el Watergate y, en particular, superior al lenguaje al que pertenece (N) que, a su vez, debería pertenecer a un lenguaje por encima de todos aquellos en los que Dean ha dicho algo sobre el Watergate... incluido (D). Situar a (D) o (N) en la jerarquía supone, pues, un círculo vicioso que no tiene nada que ver con las paradojas reflexivas.[28]

Por todos estos motivos, la teoría de Tarski resulta insatisfactoria y demasiado restrictiva como solución al problema de la paradoja del mentiroso.[29] Muchos son (Priest y Simmons entre otros) los que van más

[28] Estas objeciones se encuentran en Kripke, 1975 (pp. 57-60), y son aceptadas y citadas virtualmente por todos aquellos que escriben sobre la materia desde 1975: Haack, 1978 (pp. 167-8); Gupta, 1982 (p. 203); Barwise y Etchemendy, 1987 (pp. 6-7); Priest, *IC* (pp. 23-4); Kirkham, 1992 (281-2); Simmons, 1993 (p. 6; 142-45); Soames, 1999 (pp. 156-8); ...

[29] Incluso el diagnóstico que hace Tarski al culpar de las paradojas a los lenguajes semánticamente cerrados está equivocado. Gupta (1982, pp. 184-91) demuestra que es posible construir (con una lógica clásica) una teoría de la verdad consistente y

allá y rechazan en general cualquier solución "jerárquica" a las paradojas reflexivas, especialmente en lenguajes naturales. El carácter *ad hoc* y poco intuitivo de la maniobra unido a las restricciones expresivas que conlleva, hacen que el precio que hemos de pagar por la consistencia sea demasiado elevado.

4.5.2 Lógicas no clásicas, fallos semánticos y puntos fijos

Una manera de evitar la "jerarquización sintáctica" de términos y lenguajes que acabamos de describir es considerar el argumento de las paradojas, no como una prueba por *reductio* de la imposibilidad de definir en *L* "verdadero en *L*", sino como una prueba de la inviabilidad de una evaluación bivalente de todas las oraciones de nuestro lenguaje. Muchos han optado por cambiar de lógica y adoptar marcos semánticos que contemplen más de dos posibilidades a la hora de evaluar una oración. Obviamente, una postura tan radical necesita una justificación filosófica que no sea *ad hoc*. Los partidarios de esta salida (históricamente sucesora de las soluciones "jerárquico-sintácticas"), basan su postura en una visión "desentrecomilladora" de la verdad inspirada en PEV. Evaluar una oración que declara verdadera (V) a otra es, fundamentalmente, una cuestión de "eliminar" el predicado 'V' mediante aplicaciones sucesivas de reglas implícitas en PEV: V('p') $\Rightarrow p$, hasta alcanzar una oración que pueda ser evaluada con independencia de PEV. El problema de oraciones como μ = 'μ es falsa' o ν = 'ν es

materialmente adecuada para lenguajes que contienen su propio predicado de verdad y cierto grado de autorreferencia (insuficiente, eso sí, para expresar oraciones del mentiroso).

verdadera' es que, en ellas, por mucho que apliquemos técnicas "desentrecomilladoras", los predicados 'verdadero' y 'falso' son *ineliminables*. Éste es el diagnóstico esencial del problema, la razón por la que ni μ ni ν pueden ser verdaderas o falsas. Después, cada autor saca sus propias consecuencias de este hecho. La ineliminabilidad de los predicados semánticos en una oración tiene varias lecturas. (1) Para unos conlleva un fallo semántico o pragmático de *presuposiciones* básicas: en ocasiones, la evaluación de una oración depende de (presupone) la verdad de otras oraciones, pero paradójicamente μ y ν se presuponen a sí mismas (van Fraassen, 1968, 1970a,b); (2) otros defienden que atribuirle verdad a oraciones como μ y ν constituye un error *categorial* (Martin, 1970; Martin y Woodruff, 1975); (3) algunos sostienen que las reglas que rigen la introducción y eliminación de 'verdadero' (implícitas en PEV) *no fundamentan* la evaluación de μ y ν como verdadero o falso (Kripke, 1975); (4) para muchos el problema de fundamentación asimila en cierto sentido la semántica del predicado 'verdadero' a la de predicados *vagos* (McGee, 1989, 1991; Tappenden 1993; Soames, 1999); (5) algunos directamente afirman (o sugieren) que μ y ν *carecen de significado* o *no expresan* proposiciones (Skyrms, 1970a,b; Jc Beall, 2001, entre otros) –Skyrms llega a bloquear la ley de intercambiabilidad *salva veritate* de los idénticos en μ = 'μ es verdadera'–.[30]

[30] (En Sainsbury 1995, pp. 107-26, se hace un examen crítico de algunas de estas propuestas y de otras parecidas.) Hay aún más propuestas que parten de la ineliminabilidad de 'verdadero' en μ y ν. La más curiosa quizá sea la de Zadeh (1975), que se enfrenta al problema mediante la semántica de una lógica difusa [*fuzzy*] con múltiples valores de verdad. Dejaremos fuera esta opción por razones de espacio, sin embargo, una versión de la objeción central que aquí formularemos al resto de tentativas

Muchas de estas justificaciones son, cuando menos, polémicas. La afirmación (2), por ejemplo, de que hay un error categorial en μ o en ν al aplicar el predicado 'verdadero' parece incorrecta. Después de todo, las predicaciones afectan a oraciones, proposiciones, enunciados u otros portadores clásicos de la verdad. En el caso de (5) no está nada claro que μ, la oración del mentiroso, no sea significativa o no exprese una proposición. Esta maniobra parece *ad hoc*. Si μ no fuera significativa, *no podríamos razonar* sobre sus posibles valores de verdad llegando a la conclusión de que es paradójica. De hecho, la mayoría de paradojas descansa sobre premisas empíricas (por ejemplo: μ = 'μ es falsa', Tarski (1933, p.158) ya lo señala), lo que hace que una oración pueda ser paradójica en un contexto pero no en otro.[31] Sería absurdo negar que 'Todos los cretenses mienten siempre' es una oración significativa. Sólo reflexionando sobre el *significado* de una oración *en un contexto dado* podemos llegar a establecer si es paradójica o no, ésta es una de las enseñanzas de la objeción de Kripke antes examinada: no podemos identificar sintáctica o semánticamente entidades forzosamente paradójicas. Quizá sería más correcto afirmar que ciertas oraciones *siempre* tienen *significado$_1$* (i.e., reglas semántico-sintácticas asociadas que indican cómo determinar su valor de verdad en algunos contextos de

---

(el mentiroso reforzado) se aplica también al caso de Zadeh (véase por ejemplo, Simmons 1993, pp. 55-8). También por cuestiones de espacio, dejaremos fuera teorías que, por razones independientes (pero en parte también por las paradojas semánticas) optan por concepciones deflacionistas de la verdad: Ramsey (1927), Strawson (1950), Grover, Camp y Belnap (1975) o Horwich (1998).

[31] Veremos en 4.5.3 como los contextualistas defienden plausiblemente que hay contextos en los que incluso μ y ν reciben un valor de verdad clásico (son verdaderas o falsas) de forma no arbitraria.

uso), pero *sólo a veces* (según el contexto) *significado$_2$* (i.e., un "contenido" o proposición asociada, o expresada, con valor de verdad). Ello podría deberse, por ejemplo, al "mal funcionamiento" de ciertas oraciones en algunos contextos (Smiley, 1993). Esta idea es más razonable, pero se enfrenta al problema, quizá irresoluble (aunque común a otras interpretaciones), de determinar en *qué* contextos es paradójica una oración. En cualquier caso, no es obvio, incluso en contextos paradójicos, que μ (o ν) no exprese nada. μ dice que *cierta oración* (fácilmente identificable: *μ*) *no es verdadera*. Es cuando menos polémico y *ad hoc* afirmar que *eso* no sea decir nada. Las ideas de Kripke (3) y van Fraassen (1) son más intuitivas y satisfactorias. Ciertamente, en casos problemáticos las reglas semánticas relevantes no justifican la atribución de un valor de verdad a μ o a ν y parece claro que la verdad de μ o ν se presupone a sí misma. Estos hechos pueden tal vez justificar, como afirman McGee, Tappenden o Soames (4), que 'verdadero' reciba un tratamiento semántico similar al de los predicados vagos.[32]

No pondremos en duda los méritos de algunas de estas teorías a la hora de justificar semánticas trivalentes (la elección dominante). Mucho de lo que dicen es seguramente acertado. Lo que sí cuestionaremos, no obstante, es que nos permitan librarnos de las paradojas. Todas estas teorías, (1)-(5), tropiezan invariablemente con el mismo escollo: la paradoja del mentiroso reforzado, un hueso duro de roer. Tenemos

[32] Se pueden encontrar más objeciones a (5) en Gupta y Belnap 1993 (pp. 8-10) y en Armour-Garb 2001; objeciones a (4) –a McGee, en realidad– en Yablo 1989 y Simmons 1993; objeciones a (1) en Kearns 1970. La teoría de Kripke es ampliamente comentada por la mayoría de obras citadas.

versiones específicas: frente a (1) y (3), la oración que dice 'Esta oración no es verdadera'; frente a (2), 'Ésta no es una oración verdadera gramaticalmente bien construida'; frente a (4), 'Esta oración no es verdadera de forma definida'; frente a (5), 'Esta oración no expresa una proposición verdadera'. Tenemos también una versión general para cualquier solución que apele a una evaluación semántica polivalente (por ejemplo, trivalente): 'O esta oración no es verdadera o recibe una de las otras evaluaciones disponibles (e incompatibles entre sí)' –en realidad, la oración 'Esta oración no es verdadera' es un contraejemplo para cualquier solución polivalente, ya que, si es verdadera, no es verdadera; y si es cualquier otra cosa, entonces *no es verdadera*, lo que la convierte en verdadera. (Existen muchas otras paradojas reforzadas para (1)-(5), éstas sólo constituyen una muestra.) Cuando reflexionamos sobre una de estas oraciones desde el "marco semántico" elegido, obtenemos una contradicción especialmente diseñada para salvar los "mecanismos de control" que dicho marco genera. Frecuentemente, la oración paradójica hace explícitos esos mecanismos en su formulación construyendo por diagonalización un objeto que, si encaja en el nuevo marco semántico, entonces lo transgrede; pero si lo transgrede, entonces *cumple* lo que dice y es, por tanto, verdadera dentro de los parámetros propuestos. Así pues, estas oraciones no pueden recibir una evaluación unívoca y no contradictoria. Es verdad que, para que la paradoja "cuaje", hemos de aceptar una noción de verdad que respete una u otra versión no clásica, pero intuitivamente aceptable, de PEV, i.e., una versión que respete su carácter "desentrecomillador". Sin embargo, esta condición debería cumplirse siempre, ya que fue el carácter desentrecomillador de la verdad *lo que motivó* en primera instancia las explicaciones aducidas en (1)-(5), donde se apelaba a la *ineliminabilidad* de 'verdadero' en μ y ν.[33]

Algunas de estas soluciones a la paradoja del mentiroso obtienen, sin embargo, grandes logros. Cabe destacar especialmente la de Kripke, sin duda la más influyente después de la tarskiana. Kripke (1975) fue el primero en demostrar que es posible ofrecer una teoría de la verdad consistente y materialmente adecuada para un lenguaje que contiene su propio predicado de verdad y en el que es posible la autorreferencia.[34] Estrictamente hablando, la solución kripkeana no es clásica –véase, no obstante, Kripke 1975 (p. 64, nota 18)–, ya que las oraciones en las que los predicados semánticos no son eliminables no reciben ningún valor de verdad, su evaluación en términos clásicos *no está fundamentada*. Pese a ello, una virtud de la teoría es que, siempre que nos centremos en oraciones *fundamentadas*: (1) los operadores lógicos elegidos se comportan como los clásicos y (2) se preservan los teoremas de la lógica clásica.[35]

---

[33] Las distintas versiones del mentiroso reforzado constituyen una objeción clásica en la literatura, todos los autores citados (o casi todos) ofrecen esta objeción frente a teorías rivales y consideran (de forma insatisfactoria para sus rivales) sus efectos con respecto a su propia teoría. En Simmons 1993, capítulos 3 y 4, se puede encontrar un excelente compendio de paradojas reforzadas construidas por diagonalización a partir del marco semántico ofrecido por distintas teorías.

[34] Martin y Woodruff (1975) obtienen un resultado similar aplicando técnicas parecidas (en particular, la idea de "punto fijo" de una función). La solución formal de Kripke es, sin embargo, más completa y rica en recursos y, sobre todo, goza de una motivación filosófica mucho más sólida.

[35] Kripke considera varios modelos no clásicos: desde las superevaluaciones de van Fraassen (1970a) a las tablas de verdad trivalentes que ofrece Kleene (1952, p. 334) y habitualmente conocidas como Kleene débiles y Kleene fuertes. En su exposición

También Kripke construye una *jerarquía transfinita*, pero no de lenguajes, sino de *interpretaciones* del predicado 'V' ('verdadero') en un *único* lenguaje *L*. Cada interpretación, I_α = $<V^+_\alpha, V^-_\alpha>$, determina la extensión, V^+_α, y la antiextensión, V^-_α, del predicado 'V' en el estadio $\alpha \in$ On de la jerarquía interpretativa (V^+_α y V^-_α son disjuntos pero no agotan todas las oraciones de *L*, sólo contienen oraciones fundamentadas –i.e., preservan la consistencia pero no la completud). La idea de Kripke es reflejar, partiendo de las reglas que nos permiten introducir y eliminar 'V' de acuerdo con PEV, los cálculos mecánicos que llevamos a cabo para asignar valores de verdad a oraciones de complejidad creciente que contienen a 'V'. Kripke parte de una interpretación, I_0, que no asigna ninguna oración a la extensión o a la antiextensión de 'V' ($V^+_0 = V^-_0 = \varnothing$): $I_0 = <\varnothing, \varnothing>$. En I_0 ninguna oración que contenga a 'V' es evaluada. La siguiente interpretación, I_1, incluye en V^+_1/V^-_1 aquellas oraciones verdaderas/falsas (atómicas o compuestas por conectivas) que no contienen el predicado 'V' u otros relacionados (i.e., 'falso', etc.). A partir de aquí, se cumple en general para todo $\beta < \alpha$ que, si *p* es verdadera en I_β, entonces la oración que dice que *p es verdadera*: V('*p*'), es verdadera en I_α. El paso de una interpretación I_β (o de un conjunto de interpretaciones) a su sucesora I_α (o a la interpretación límite inmediatamente superior en la jerarquía), puede describirse mediante una función, *s*, que "salta" de una(s) a otra *declarando* como verdadero en I_α

Kripke opta por las últimas, pero afirma que la elección entre los modelos que cita debería ser irrelevante desde el punto de vista de los resultados lógicos (1975, p. 76). Así mismo, ofrece una definición no clásica de los cuantificadores.

todo aquello que *es* verdadero en I_β (s es un generador transfinito: $s(\{I_\beta: \beta < \alpha\}) = I_\alpha$, donde α es el ordinal sucesor de β o el límite de una serie infinita de βs). Debido al modo en que hemos definido las interpretaciones y las conectivas lógicas, Kripke demuestra que s tiene ciertas propiedades formales. En primer lugar, s es un operador monótono: $I_\beta \subseteq I_\alpha, \rightarrow s(I_\beta) \subseteq s(I_\alpha)$; en segundo lugar, $I_\beta \subseteq I_\alpha$ (para todo $\beta < \alpha$). Intuitivamente esto significa que todo lo que es interpretado como verdadero o falso conserva su valor de verdad en subsiguientes interpretaciones. Lo único que cambia es que algunas oraciones que no tenían ningún valor de verdad en una interpretación determinada reciben uno en alguna interpretación posterior (y ya no lo perderán). Lo curioso de esta jerarquía interpretativa es que, gracias a la monotonía de s, se puede demostrar que existe una interpretación I_α tal que $s(I_\alpha) = I_\alpha$ (s tiene un *punto fijo mínimo*). Esto quiere decir que I_α es una interpretación a partir de la cual no se pueden evaluar como verdaderas o falsas nuevas oraciones. Todo lo que es declarado verdadero (falso) en I_α, es verdadero (falso) en I_α y viceversa. En otras palabras, I_α satisface *todas* las ejemplificaciones de PEV: V('*p*') $\leftrightarrow$ *p*. Kripke muestra además que en I_α ninguna oración paradójica o no fundamentada recibe valor de verdad alguno, así pues, I_α proporciona una teoría de la verdad consistente y materialmente adecuada.[36]

[36] Además del mínimo, s tiene otros puntos fijos, Kripke los utiliza para definir "oración paradójica" frente a "no fundamentada". Una breve exposición de la teoría kripkeana se puede encontrar en Kirkham 1992 (pp. 282-94); para una exposición crítica y más detallada en el terreno formal véase Simmons 1993 (pp. 47-55). Soames 1999 (capítulo 6, especialmente, pp. 181-200) contiene una de las exposiciones más completas.

¿Cómo soluciona Kripke el problema del mentiroso reforzado? En la teoría de Kripke, μ = '$\neg V(\mu)$' es una oración no fundamentada y, por tanto, *ni verdadera ni falsa* en I_α, pero como μ dice de sí misma que no es verdadera, lo que dice –i.e., $\neg V(\mu)$– *es verdad* y tenemos una paradoja. La manera que tiene Kripke de bloquear esta paradoja es decir que, dada su definición de '$\neg$', si μ no tienen valor de verdad, tampoco '$V(\mu)$' y '$\neg V(\mu)$' lo tienen. Para expresar la paradoja, necesitaríamos una definición alternativa de '$\neg$' según la cual si p no es verdadera (i.e., es falsa o no fundamentada), entonces $\neg\ p$ es verdadera. Esta noción se puede expresar en los lenguajes naturales, pero *no* en el lenguaje L de Kripke. Kripke (1975, pp. 79-81) reconoce esta restricción y admite que, en última instancia, el problema sólo podría solucionarse parcialmente apelando a una jerarquía tarskiana de lenguajes de creciente complejidad semántica. Pero ya conocemos (en parte por Kripke y, en parte por 4.5.1) qué problemas deben afrontar estas jerarquías semántico-sintácticas.

Pese a todo, es innegable que la solución de Kripke representa un avance en todos los sentidos con respecto a la teoría tarskiana. Las técnicas de punto fijo de Kripke y Martin y Woodruff (1975) han sido desarrolladas y ampliadas por Gupta (1982), Herzerberger (1982), Feferman (1984), Barwise y Etchemendy (1987), Visser (1989), McGee (1991), Gupta y Belnap (1993) y un sinfín de autores.[37] Lo que muestra, por otra parte, que la teoría de Kripke no es del todo satisfactoria y ello,

[37] En Barba (1998), encontramos un buen compendio de técnicas diferentes para alcanzar puntos fijos en la interpretación de lenguajes con predicados semánticos como 'verdadero'. Visser (1989 pp. 647 y siguientes), Gupta y Belnap (1993, esp. capítulo 2)

en gran medida, por la paradoja del mentiroso reforzado. Tampoco los autores que han secundado técnicas de punto fijo en su solución a las paradojas se han librado de las paradojas reforzadas.

4.5.3 Teorías contextualistas: verdadero en un contexto

Para la mayoría de autores, ninguna solución a las paradojas que soslaye o ignore la versión clásica del mentiroso reforzado, μ = 'μ no es verdadera', puede ser satisfactoria. Burge (1979) fue uno de los primeros en sugerir que μ sólo supone un problema cuando olvidamos que las oraciones *sólo* pueden recibir un valor de verdad si son evaluadas *en un contexto* y que una misma oración puede recibir *distintos* valores de verdad en contextos diferentes (i.e., puede decir *distintas* cosas).[38] Según Burge, ése es el problema que μ plantea. Cuando intentamos evaluar μ desde su contexto de proferencia, *c*, nos damos cuenta mediante un razonamiento, r_1, de que μ no puede ser verdadera (ni falsa) coherentemente. Así pues, concluimos que *μ no es verdadera*. Ahora bien, esta conclusión se usa como premisa en un nuevo razonamiento, r_2, *que tiene lugar en un contexto diferente*, *c'*, para concluir que *μ es verdadera*. La paradoja se resuelve señalando que μ puede ser verdadera en *c'* aunque no lo sea en *c*, r_1 establece lo segundo: μ no es verdadera *en c*; y r_2 lo primero: μ es verdadera *en c'*.[39]

---

y Smullyan (1994) ofrecen quizá las exposiciones más detalladas de este tipo de técnicas.

[38] Anticipándose a Burge, Parsons (1974) es el primero en apuntar una solución contextualista.

[39] Desde la pragmática, Burge (1979, pp. 93-9) explica ingeniosamente el modo en que se genera la paradoja apelando a la presencia de una implicatura: a saber, el supuesto

La línea argumental de Burge ha sido adoptada, entre otros, por Barwise y Etchemendy (1987), Gaifman (1992) y Simmons (1993). Una de las ventajas de las teorías contextualistas es que, al considerar como portadores de verdad, no a las oraciones [*type*], sino a las proposiciones que éstas expresan *en un contexto de uso* (o a los enunciados, o a las oraciones [*token*]), pueden operar con lógicas clásicas incluso al hablar de oraciones como $\mu_{[type]}$. Aunque esto no elimina el problema de fondo (al menos para Burge, 1979, p. 96), aún existen contextos *c* paradójicos en los que la evaluación de una oración $\mu_{[token]}$ no es clásica (i.e., $\mu_{[token]}$ no es ni verdadera ni falsa en *c*) y requiere, a menudo, una lógica diferente. Con todo, los contextualistas hacen más justicia que el propio Kripke a la idea de que las paradojas no se pueden aislar apelando a criterios sintácticos: $\mu_{[type]}$ no es una oración no fundamentada, sólo ciertos *usos* de $\mu_{[type]}$ involucran problemas de fundamentación, otros son perfectamente válidos.

Por desgracia, todas estas teorías (obviamente paramétricas) tropiezan con un problema común, su propia versión reforzada del mentiroso: el mentiroso contextual o "super mentiroso", cuya expresión más popular es μ = 'μ no es verdadera *en ningún contexto de proferencia*'. Si μ es verdadera en algún contexto, entonces es falsa en *ese* contexto, si no es verdadera en ningún contexto, dado que eso es lo que μ afirma, entonces debe ser verdadera en *alguno*. Todas las teorías

conversacional de que μ se evalúa *en ambos casos*, r_1 y r_2, desde *c*. Si cancelamos dicha implicatura en el caso de r_2, la paradoja se desvanece: r_2 concluye que μ es verdadera *en c'* porque r_1 concluye que μ no es verdadera *en c*.

contextualistas son conscientes de esta paradoja y la afrontan, al igual que las teorías examinadas anteriormente, introduciendo las restricciones necesarias para que su versión particular de la paradoja sea *inexpresable*.

Burge, por ejemplo, afirma que aunque nuestro lenguaje contiene un único predicado 'verdadero', éste se comporta como los demostrativos y otros índices: su extensión cambia según el contexto (1979, pp. 94; 107). Burge introduce una jerarquía de contextos (de hecho, varios modelos opcionales) representados por índices 'i' que acompañan al predicado 'verdadero' (verdadero$_i$). El propósito de la jerarquía no es restringir el ámbito de aplicación de 'verdadero', sino definir en *qué* contextos de reflexión podría ser patológica la emisión de una oración. Sin embargo, Burge se limita a desautorizar la posibilidad de hablar o cuantificar *en general* sobre los *índices* que afectan al predicado verdadero (i es siempre un índice *oculto*). Así pues, evita la paradoja contextualista reforzada prohibiendo que hagamos explícitos en nuestro discurso los factores que permiten su expresión. Esta jugada no es convincente, entre otras cosas, porque el propio Burge ha hecho explícitos esos elementos (quebrantando su propia prohibición) a la hora de ofrecernos su teoría de la verdad en una suerte de metalenguaje informal. Es más, Burge ha hecho afirmaciones *generales* sobre la jerarquía, cuantificando informalmente sobre los índices contextuales.[40]

---

[40] Otra dificultad para Burge es explicar cómo podemos determinar dentro de la jerarquía en *qué* contexto cabe evaluar una proferencia. Burge se enfrenta al problema ofreciendo varias máximas pragmáticas y diferentes modelos jerárquicos. La complejidad de su propuesta muestra la magnitud del problema. Se puede encontrar otras críticas a Burge (acompañadas a veces de exposiciones de su teoría) en Gupta

Barwise y Etchemendy (1987) "resuelven" la paradoja del mentiroso mediante la confección de una compleja semántica situacional elaborada a partir de las siguientes nociones:[41]

(1) *Estados de cosas* (*op. cit.* pp. 75-6): Se representan mediante n+2-tuplas: <P, x_1, ..., x_n; *i*>, en las que P es una relación n-aria; x_1, ..., x_n son objetos; e *i*, un número: 0 ó 1, que señala si P *se da* entre x_1, ..., x_n (1); o *no se da* (0). Ejemplo: <Correr, *Juan*; 0> representa un estado de cosas negativo en el que Juan no corre. Todo estado de cosas positivo, $<e^1>$, tiene un *dual* negativo, $<e^0>$ y viceversa.

(2) *Situación* y *Mundo* (pp. 75-6; 131-2): Una *situación* es un conjunto, *s*, de estados de cosas, i.e., *s* = {<P, *a*; 1>, <Q, *a*, *b*; 0>, ...}. Un "*Mundo*", *M*, es un conjunto *completo* de estados de cosas, $<e^i>$, consistentes entre sí. (i.e., *M* satisface ciertos requisitos de coherencia y completud: para *todo* $<e^i>$, o $<e^1> \in M$ o $<e^0> \in M$). Cualquier subconjunto de *M* (incluido *M*) *es una situación* (las situaciones también han de ser coherentes).

(3) *Proposición "austiniana"* (pp. 28-30, 121-4):[42] Es aquello que *expresa* una oración en un contexto de uso, *c*. Toda proposición (*p*) *versa sobre una situación* (identificable mediante convenciones demostrativas) *y sobre un tipo de estado de cosas* (identificable mediante convenciones descriptivas). Ésos son sus constituyentes esenciales. Toda proposición

1982 (pp. 203-4), Sainsbury 1995 (p., 126), Gupta y Belnap 1993 (pp. 11-2) o Simmons 1993 (pp. 96-8)

[41] Barwise y Etchemendy (*op. cit.* pp. 61-118) ofrecen también una teoría *no* situacional que omitiremos aquí. Un excelente resumen de su teoría situacional se puede encontrar en Kirkham 1992, pp. 298-306.

(p) *dice* de una situación, s, que es de cierto tipo, $[e^i]$ (o, para ser más precisos, (p) *dice* que un estado de cosas $<e^i>$ del tipo $[e^i]$ se da en s: $<e^i> \in s$). Una proposición se representa mediante un conjunto cuyos componentes son la situación, s, y el *tipo de estado de cosas*, $[\mathrm{P}, x_0, ..., x_n; i]$, sobre los que versa. Por ejemplo, si emitimos en un contexto c la oración ($\underline{p}$): $\underline{\mathrm{Correr}(\mathit{Juan})}$,[43] expresamos la proposición (j): $\{s; [\mathrm{Correr}, \mathit{Juan}; 1]\}$, cuyo contenido es el siguiente: "La situación s es del tipo en el que Juan corre (i.e., *Juan corre en s* o, formalmente, $<\mathrm{Correr}, \mathit{Juan}; 1> \in s$)". Las proposiciones pueden figurar como objetos en estados de cosas, además, una proposición puede hablar de *sí misma*.

(4) *Verdad* y *falsedad* (pp. 126-8): Una proposición (p): $\{s; [\mathrm{P}, x_0, ..., x_n; i]\}$ es *verdadera* si, y sólo si, $<\mathrm{P}, x_0, ..., x_n; i> \in s$ ($i \in \{0, 1\}$) y *falsa* en caso contrario (i.e., $<\mathrm{P}, x_0, ..., x_n; i> \notin s$). "Ser falso" es equivalente, pues, a "no ser verdadero". Sin embargo, $<\mathrm{P}, x_0, ..., x_n; i> \notin s$ *no implica* $<\mathrm{P}, x_0, ..., x_n; i^{\neg}> \in s$ ($i \in \{0, 1\}$ & $i \neq i^{\neg}$). Esto es, la falsedad de una proposición $\{s; [e^i]\}$ *no implica* la verdad de su dual $\{s; [e^{i\neg}]\}$. Una situación, s, del tipo en el que Juan no corre (i.e., $<\mathrm{Correr}, \mathit{Juan}; 0> \in s$) es una situación en la que la proposición (j): $\{s; [\mathrm{Correr}, \mathit{Juan}; 1]\}$ es falsa y su dual ($j^{\neg}$): $\{s; [\mathrm{Correr}, \mathit{Juan}; 0]\}$ verdadera. Una situación, s', que *no aporta información* con respecto a si Juan corre o no (i.e., $<\mathrm{Correr}, \mathit{Juan}; 1> \notin s'$ y $<\mathrm{Correr}, \mathit{Juan}; 0> \notin s'$) es una situación en la

---

[42] Barwise y Etchemendy ofrecen un modelo formal para la noción de proposición desarrollada por Austin en 1950.

[43] En $\underline{\mathrm{P}}(\underline{a})$, el subrayado señala que $\underline{\mathrm{P}}$ y $\underline{a}$ son expresiones lingüísticas (un predicado y un nombre en este caso) que representan respectivamente a P (una propiedad o relación) y a a (un objeto).

que $(\bar{j})$ y (j) son *ambas* falsas, ya que *s*' no es ni de tipo [Correr, *Juan*; 1] ni de tipo [Correr, *Juan*; 0].

(5) *Negación$_1$* y *negación$_2$* (pp. 16-8; 164-70). La definición austiniana de proposición propicia que negar una oración como $\underline{\mathrm{P}}(\underline{a})$ puede significar dos cosas distintas: $\underline{\mathrm{no}_1}$ $\underline{\mathrm{P}}(\underline{a})$ [*Negation*]: "La situación *s* relevante es del tipo en que *a no es* P", i.e., (p_1): $\{s;\ [\mathrm{P}, a;\ 0]\}$; o $\underline{\mathrm{no}_2}$ $\underline{\mathrm{P}}(\underline{a})$ [*Denial*]: "La situación *s* relevante *no es* del tipo en que *a* es P", i.e., (p_2): $\neg\{s;\ [\mathrm{P}, a;\ 1]\}$. La proposición (p_1) es verdadera si y sólo si $<\mathrm{P}, a;\ 0> \in s$, mientras que la proposición (p_2) es verdadera si y sólo si $<\mathrm{P}, a;\ 1> \notin s$. Dada una situación *s*, no se cumple en general que $<\mathrm{P}, x_0, ..., x_n;\ i> \in s \leftrightarrow <\mathrm{P}, x_0, ..., x_n;\ \bar{i}> \notin s$ (vimos en (4) que falla '$\leftarrow$').

La solución que se ofrece a la paradoja del mentiroso descansa precisamente en la idea de que la negación es *ambigua*. El razonamiento que establece la conclusión paradójica confunde erróneamente los dos sentidos en que se puede negar una oración (i.e., las dos proposiciones diferentes que una oración negada ambigua puede expresar). Pero, antes de ver la solución en detalle, cabe destacar que, a partir de los conceptos descritos, Barwise y Etchemendy construyen una teoría, $\mathrm{T_{BE}}$, con las siguientes carácterísticas:

(I) $\mathrm{T_{BE}}$ respeta una versión de PEV para proposiciones. Dada una proposición cualquiera (p): $\{s;\ [e^i]\}$, se cumple siempre que: la situación *s* es de tipo $[e^i]$ si y sólo si $<e^i> \in s$ (prop. 1.2, p. 127). Esto permite las siguientes equivalencias (prop. 2.2, p. 127): $\{s;\ [e^i]\}$ es verdadera $\leftrightarrow$ *s* es de tipo $[e^i] \leftrightarrow <e^i> \in s$. $\mathrm{T_{BE}}$ es, por tanto, *materialmente adecuada*.

(II) Las *proposiciones* de T_{BE} son *bivalentes* (clásicas): Para toda proposición (p): $\{s; [e^i]\}$, se cumple que (prop. 1.1, p. 127) o s es del tipo $[e^i]$ o no lo es (i.e., o $<e^i> \in s$ o $<e^i> \notin s$). Y, por tanto, se cumple que (prop. 2.1, p. 127) (p) es verdadera o falsa (i.e., $<V, p; i> \in s$ o $<V, p; i> \notin s$).

(III) Además de las características ya señaladas en (2), un modelo total (un "Mundo"), M, satisface (prop. 5, p. 132) dos requisitos adicionales. Para toda proposición (p):

(IIIa) (p) es verdadera $\leftrightarrow <V, p; 1> \in M$

(IIIb) (p) es falsa $\leftrightarrow <V, p; 0> \in M$.

Llegados a este punto, nos preguntamos cómo son posibles (I)-(III) teniendo en cuenta la presencia de la paradoja del mentiroso, i.e., la proposición μ_1: $\{s; [V, \mu_1; 0]\}$.[44] T_{BE} resuelve la paradoja afirmando que μ_1 *es falsa*. El razonamiento es el siguiente (teorema 6, pp. 132-3): μ_1 no puede ser verdadera consistentemente. Si lo fuera, $<V, \mu_1; 1> \in M$ (por IIIa) y $<V, \mu_1; 0> \in s \subseteq M$ (por I), contradiciendo el requisito de coherencia que (2) impone a M. Ahora bien, si μ_1 no puede ser verdadera, *ha de ser falsa* (por II).

[44] La paradoja del mentiroso entraña aquí una dificultad añadida. En vista de la existencia de dos modelos de negación, la oración del mentiroso, μ: $\underline{\text{no}}\ \underline{V}(\underline{\mu})$, puede expresar dos tipos de proposición según contextos: μ_1: $\{s; [V, \mu_1; 0]\}$, o μ_2: $\neg\{s; [V, \mu_2; 1]\}$. Ambas dicen de sí mismas que son falsas en s (ya que "ser falso" $\equiv$ "no ser verdadero"), pero lo que dicen es diferente: μ_1 dice que *su falsedad se da* en s ($<V, \mu_1; 0> \in s$); μ_2 dice que *su verdad no se da* en s ($<V, \mu_2; 1> \notin s$). T_{BE} muestra que las proposiciones de tipo μ_1 son falsas y las de tipo μ_2 verdaderas. Nos centraremos en la solución a μ_1. Para ver μ_2 consúltese Barwise y Etchemendy, 1987 (pp. 164-70) o Kirkham, 1992 (pp. 303-4). La solución es similar a la de μ_1.

La paradoja surge cuando razonamos del siguiente modo: "Si μ_1 es falsa, i.e., $<V, \mu_1; 0> \notin s$, *entonces su dual es verdadera*, i.e., $<V, \mu_1; 1> \in s$, luego $<V, \mu_1; 1> \in M$ ($s \subseteq M$). Pero esto significa (por IIIa) que μ_1 es verdadera, contradiciendo que es falsa". El error de este razonamiento es fácil de identificar. Según (4) y (5), $<V, \mu_1; 0> \notin s$ *no implica* $<V, \mu_1; 1> \in s$, por tanto de la falsedad de μ_1 *no se sigue* su verdad. Podemos entender, no obstante, por qué se produce este error. Al razonar nos decimos "μ_1 dice de sí misma que es falsa y, de hecho, μ_1 *es falsa*, luego lo que dice es verdad y, por PEV, μ_1 *es verdadera*." Nuestro error radica en que hemos identificado mal *lo que dice* μ_1. Ciertamente, μ_1 es falsa, *pero eso no es lo que dice* μ_1. Lo que dice μ_1 es "la falsedad de μ_1 *se da* en s" (i.e., "$<V, \mu_1; 0> \in s$"). Según (4), μ_1 es falsa si, y sólo si, "la falsedad de μ_1 *no se da* en s" (i.e., "$<V, \mu_1; 0> \notin s$"), pero ése es el contenido de la proposición μ_1*: $\neg\{s; [V, \mu_1; 0]\}$, y (5) deja bien claro que $\mu_1^* \neq \mu_1$. Hemos visto que lo que dice μ_1, i.e., "$<V, \mu_1; 0> \in s$" no implica en modo alguno lo que dice μ_1*, i.e., "$<V, \mu_1; 0> \notin s$". Así pues, nada de lo que dice μ_1 nos lleva a afirmar "μ_1 es verdadera". μ_1 es falsa sin más, lo que dice *no se cumple*.

Lo más curioso es que (por IIIb) si μ_1 es falsa, entonces $<V, \mu_1; 0> \in M$ y, por tanto, *debe existir* un s' tal que $<V, \mu_1; 0> \in s' \subseteq M$. Sin embargo, no hay aquí ninguna incoherencia, sino un reflejo de intuiciones contextualistas: μ_1 es falsa porque, en contra de lo que afirma, su falsedad *no* se da en s (i.e., $<V, \mu_1; 0> \notin s$), pero existe una situación s' *diferente* de s (v.g., $s' = M - s$) tal que la falsedad de μ_1 *se da en* s'. Hay, pues, proposiciones verdaderas (como μ_1*) que expresan la falsedad de μ_1, pero esas proposiciones expresan una falsedad que no se da en *s*, sino en un contexto distinto de aquel sobre el que versa μ_1. Esto es posible por el siguiente hecho:

(IV) Según (2), todo "Mundo", *M*, es tal que (para *todo* $<e^i>$) o $<e^1> \in M$ o $<e^0> \in M$. En T_{BE}, existen situaciones $s \neq M$ en las que (para *algún* $<e^i>$) $<e^1> \notin s$ y $<e^0> \notin s$. En estos casos, una situación no es ni del tipo $[e^1]$ ni del tipo dual $[e^0]$. Esto indica que las situaciones $s \neq M$ *no son completas* en el sentido en que *sí* lo es el Mundo, *M*, en que se incluyen. Las proposiciones paradójicas, μ, son siempre de este tipo, versan sobre situaciones, *s*, que no son ni del tipo enunciado por μ ni del tipo dual. Pero *siempre* existe una situación $s' \neq s$ en la que *se da* la verdad o falsedad de μ y, por tanto, μ es verdadera o falsa en *M*.

La paradoja del mentiroso, motivada por la oración $\underline{\mu}$: $\underline{\mu}$ <u>no es verdadera</u>, es el fruto de una ilusión: Al razonar oscilamos entre dos proposiciones *distintas*, μ_1: $\{s; [V, \mu_1; 0]\}$ (i.e., "*s* es una situación en la que μ_1 no es verdadera") y μ_1*: $\neg\{s; [V, \mu_1; 0]\}$ (i.e., "*s* NO es una situación en la que μ_1 no es verdadera"). Si μ_1 y μ_1* fuesen equivalentes, entonces tendríamos, obviamente, una paradoja, pero no tenemos ninguna razón para pensar en T_{BE} que μ_1 y μ_1* compartan el mismo contenido. Las dos negaciones que aparecen en μ_1*, 'NO' y 'no', son distintas y guardan relación con distintos componentes de μ_1*, a saber, '$\neg$' y '0', respectivamente. El problema de nuestro lenguaje es que la palabra 'no' significa unas veces '$\neg$' y otras el operador que cambia la polaridad de una proposición de '0' a '1' y viceversa. Al razonar sobre la oración <u>μ</u>, la paradoja surge cuando pasamos por alto la ambigüedad del término <u>no</u> y confundimos '¬' [*denial*] con el operador de polaridad [*negation*].

La solución de Barwise y Etchemendy es realmente ingeniosa y sofisticada. La semántica situacional que ofrecen refleja intuiciones interesantes sobre la ambigüedad del concepto de negación y la relevancia de las situaciones en la evaluación semántica. Otro mérito objetivo es el original y novedoso uso de conjuntos no fundamentados para construir modelos de proposiciones autorreferentes (representadas como conjuntos que aparecen en su propia clausura $\in$-transitiva –véase *op. cit.*, pp. 44 y 62).[45] Barwise y Etchemendy también usan teoremas de punto fijo en sus modelos semánticos no fundamentados, habitualmente identifican ciertos conjuntos relevantes con el punto fijo maximal de ciertas operaciones monótonas. Pero, pese a este despliegue de recursos técnicos y distinciones conceptuales, cabe hacer algunas críticas relevantes.

El problema más grave de T_{BE} hace referencia de nuevo a limitaciones expresivas severas inducidas por la presencia de paradojas reforzadas. También T_{BE} tiene sus propias paradojas situacionales reforzadas. Consideremos, por ejemplo, la siguiente proposición, μ_M: $\{M; [V, \mu_M; 0]\}$, μ_M dice de sí misma que es falsa en una situación *global*, i.e., en el *Mundo*, *M*. Por el razonamiento esgrimido en el caso de μ_1, μ_M no

[45] Sin embargo, McLarty (1993) ofrece argumentos de peso para desechar la idea de que representar estructuras autorreferentes *requiera* la participación de teorías de conjuntos no fundamentados. Con las teorías clásicas podemos cumplir con creces los mismos objetivos.

puede ser verdadera coherentemente, así pues, por (II), ha de ser falsa, i.e., <V, μ_M; 0> $\notin M$. Sin embargo, esta vez la situación es diferente porque, por (IIIb), tenemos una contradicción explícita: <V, μ_M; 0> $\in M$ y aquí no hay maniobra que valga.[46] La manera que tienen Barwise y Etchemendy de salvar esta paradoja es excluir la posibilidad de proposiciones que versen sobre situaciones totales, i.e., sobre el mundo. Frente a la suposición de que una situación s expresable es total ($s = M$) podemos imaginar siempre una proposición μ que *diagonalice* fuera de s mostrando que no puede ser total consistentemente (pp. 154-5). Si construimos una nueva situación más comprensiva, s', tendremos un nuevo mentiroso, μ', de modo que ninguna jerarquía creciente de situaciones expresables podrá formar una situación M. Pero, una vez más, esta solución es insatisfactoria por los sacrificios *ad hoc* que nos impone (¿por qué no puede haber proposiciones sobre M?) y porque hace ininteligible la propia teoría T_{BE}, que contiene numerosas proposiciones (incluso teoremas) que nos hablan del mundo, de su completud y coherencia.

Simmons (1993) diseña otra teoría contextual denominada de la "singularidad" e inspirada en soluciones contextualistas medievales y modernas a la paradoja del mentiroso (*op. cit.*, pp. 83-98).[47] Al igual que dictan sus predecesoras, una oración sólo puede ser evaluada como

[46] En el caso de μ_1 salvábamos la paradoja porque *sólo* podíamos inferir <V, μ_1; 0> $\in M$ a partir de <V, μ_1; 0> $\notin s$, <V, μ_1; 0> $\in s$', enunciados todos ellos consistentes entre sí, ya que $s \neq s$'.

[47] Simmons sigue ideas y técnicas muy similares a las de Gaifman (1992). La diferencia fundamental radica en que, a diferencia de éste, Simmons no aboga por teorías jerárquicas (como veremos).

verdadera, falsa o paradójica en un contexto de emisión determinado, por ello Simmons (pp. 101-2) se centra en oraciones [*token*], u oraciones [*type*] + contexto, que representaremos mediante $[p]_c$ ('*p*' es una oración [*type*] y '*c*' el contexto en que se ha usado). Una oración $[p]_c$ es paradójica en contextos en los que su evaluación semántica no está fundamentada (aquellos en que los predicados semánticos de $[p]_c$ no son eliminables), en el resto de contextos, $[p]_c$ es verdadera o falsa.[48] La novedad fundamental de la propuesta de Simmons radica en su rechazo de cualquier solución jerárquica a las paradojas. Simmons critica las soluciones jerárquicas sintáctico-semánticas sobre la base de que nuestros lenguajes naturales no esconden ninguna jerarquía de predicados, lenguajes, etc. (pp. 99, 159), son "universales" en el sentido tarskiano de la nota 7.

Al igual que Burge, Simmons afirma que el predicado 'verdadero' es único, pero su extensión varía según contextos (p. 101): Cada contexto, *c*, constituye un marco de reflexión diferente desde el que efectuar la evaluación de una oración como $[V([p]_c)]_{c'}$. La reflexión evaluadora se representa mediante un árbol con ramas y nodos de tipo: $[V([p]_c)]_{c'}$--$[p]_c$-- ... en el que, salvo el primero, cada nodo es el resultado o bien de la aplicación de reglas semánticas basadas en PEV o bien de la descomposición de una oración compleja en sus constituyentes (por ejemplo, de $[[A]_c \wedge [B]_{c'}\ ]_{c''}$ saldrían dos ramas cuyos primeros nodos serían respectivamente $[A]_c$ y $[B]_{c'}$). Si el árbol que genera la reflexión

[48] Al igual que Kripke, Simmons utiliza (p. 123, n. 6) el esquema trivalente "Kleene fuerte" para definir las conectivas lógicas cuando operan sobre oraciones no

evaluativa de una oración $[p]_c$ tiene alguna rama *infinita*, entonces $[p]_c$ podría no estar fundamentada en c (i.e., los conceptos semánticos de $[p]_c$ podrían no ser eliminables en c).[49] Pese a todo, aunque $[p]_c$ no estuviese fundamentada en c (o en algún otro contexto), podría haber contextos c' en los que una reflexión sobre la evaluación de $[p]_c$ generase un árbol finito o un árbol cuyas ramas infinitas no impidiesen que $[p]_c$ fuese verdadera o falsa (i.e., que estuviese fundamentada). Para reflejar este hecho, Simmons construye una "jerarquía reflexiva" (pp. 124-5) en la que cada contexto se identifica con un índice, n, que acompaña al predicado 'verdadero' (y a PEV), según el contexto evaluativo en que tiene lugar la reflexión. La primera "reflexión" evaluadora tiene lugar en el contexto 0, habitualmente el contexto de proferencia de la oración [*type*] '*p*' (en el caso de $[p]_c$, 0 sería c). A partir de aquí, y básicamente por las razones pragmáticas aducidas por Burge (véase nota 39), las oraciones evaluadas en primera instancia como no fundamentadas generan nuevos contextos de reflexión 1, 2, 3, ... y nuevas evaluaciones. Simmons afirma que, para toda oración $[p]_m$, existe siempre un contexto de reflexión, n, tal que $[p]_m$ es verdadera$_n$ o falsa$_n$.

Lo que diferencia a la teoría de Simmons de otras es que en su jerarquía no existe ningún nivel, m, a partir del cual (o por debajo del

---

fundamentadas.

[49] Que la oración esté o no fundamentada depende de varias cosas, entre otras de la definición que elijamos para las conectivas trivalentes. En Kleene fuerte, sería verdadera (estaría fundamentada) una disyunción: $[[A]_c \vee [B]_{c'}]_{c''}$, que generase, por ejemplo, un árbol con dos ramas: una *finita* cuyo primer nodo, $[A]_c$, resultase verdadero; y otra *infinita* cuyo primer nodo, $[B]_{c'}$, resultase no fundamentado. Si $[A]_c$ fuese falsa en este

cual) una oración sea siempre paradójica y no pueda ser por tanto 'verdadera$_n$' para todo n mayor o menor que m. En la jerarquía de Simmons una oración podría ser verdadera$_n$, no fundamentada$_{n+1}$ y verdadera$_{n+2}$, por ejemplo. Los árboles de Simmons determinan reflexivamente en *qué* contextos es paradójica una oración $[p]_c$ y la excluyen de la extensión de 'verdadero' (o 'falso') precisamente en *esos* contextos y no en otros. De ahí el nombre de la teoría: Si la evaluación de $[p]_c$ no está fundamentada en *c*' decimos que $[p]_c$ es una "*singularidad*", en el contexto *c*', para los predicados 'verdadero'/'falso' (i.e., no figura en su extensión en *c*'), pero eso no afecta a contextos de orden superior o inferior. Las restricciones han de ser *mínimas* para conservar al máximo la capacidad expresiva del lenguaje (pp. 106-8).

Simmons (pp. 159-82) se esfuerza por mostrar que su teoría es "universal" y puede expresar sus propios conceptos, nociones tales como *verdadero en un contexto* n (i.e., *verdadero*$_n$) o *fundamentado*. El obstáculo fundamental, una vez más, es la paradoja del mentiroso contextual (pp. 171-3): μ = '(La oración [*type*]) μ no es verdadera *en ningún contexto*'. La manera que tiene Simmons de afrontarla es muy intrincada y poco convincente. Para evitar la inexpresibilidad de la paradoja y la jerarquización del lenguaje ordinario, Simmons introduce una distinción *única* entre dos lenguajes (pp. 174-81), un lenguaje objeto (LO) con términos semánticos sensibles a variaciones contextuales y el (meta)lenguaje (ML) de la teoría, que evalúa a las oraciones como

---

ejemplo, $[[A]_c \vee [B]_{c'}]_{c''}$ no estaría fundamentada en *c*''. Cuando todas las ramas de un árbol son finitas, la oración inicial $[p]_c$ es siempre verdadera o falsa.

verdaderas o falsas con "independencia" de toda reflexión contextualizada. En ML cuantificamos sobre contextos y tenemos un predicado: 'verdadero$_{\alpha}$ para algún contexto α' ('verdadero$_{obj}$', para abreviar) *no sensible a variaciones contextuales.*[50] La idea parece ser la siguiente: Según Simmons, en su teoría *todas* las oraciones (incluida μ) son (sólo) verdaderas o (sólo) falsas *en algún contexto reflexivo* n. Si esto es así, será *verdadero en todo contexto* (y, por tanto, con independencia del contexto) decir de una oración verdadera en n *que es verdadera$_{\alpha}$ en algún contexto* α (i.e., verdadera$_{obj}$). Cuando reflexionamos sobre μ nos percatamos de que μ es paradójica en *todos* los contextos y, consiguientemente, concluimos que "no es verdadera en ninguno". Puesto que eso es lo que afirma μ, inferimos que μ es verdadera. Pero si μ es verdadera en alguna reflexión evaluativa, entonces es verdadera$_{obj}$ con independencia del contexto. Para enfatizar la idea de que hay oraciones verdaderas con independencia del contexto, Simmons compara μ con la oración h = "*'2+2 = 4' es verdadera en cualquier contexto de emisión*". Ambas, μ y h, son verdaderas con independencia del contexto (verdaderas$_{obj}$) porque (1) afirman que reciben la misma evaluación semántica en todo contexto, paradójica y verdadera, respectivamente, y (2), de hecho, la reciben.

[50] "[I]t *is* the case that the language of the theory [ML] contains a truth predicate for the object language [LO]: the extension of 'true$_{\alpha}$ for some α', or more conveniently, 'true$_{obj}$', comprises exactly the sentences of [LO] that are true from the context independent perspective. And it is the case that 'true$_{obj}$' is not a predicate of [LO]. Context independent uses of 'true' are not to be found in [LO]: In [LO], uses of 'true' and 'false' are context sensitive uses. The sentence ['μ is true$_{obj}$'] is a sentence of [ML]" (ibíd., pp. 174-5)

Tanto la solución como la analogía plantean serias dudas y dificultades de comprensión. Volvamos al razonamiento cuya conclusión era: 'μ es verdadera$_{obj}$'. Sin duda, este razonamiento *tiene lugar en un contexto*, al que llamaremos R_μ. En R_μ nos percatamos de que, en *todo* contexto, μ es verdadera si y sólo si no es verdadera, así pues, concluimos que μ no está fundamentada en *ningún* contexto y, por tanto, que no es verdadera en *ningún* contexto. Ahora bien, como *eso* es lo que afirma μ, inferimos (en R_μ) que μ es verdadera$_{R\mu}$ y, como μ es verdadera en al menos un contexto, i.e. en R_μ, μ es verdadera$_{obj}$. La cuestión crucial es: *¿Forma parte* R_μ *de los contextos sobre los que habla* μ *y en los que la evaluamos o no*? Si R_μ *forma* parte de esos contextos, entonces si μ fuese verdadera$_{R\mu}$, no sería verdadera$_{R\mu}$ (por $PEV_{R\mu}$) y viceversa: μ sería verdadera y no verdadera (paradójica) en R_μ, al igual que en el resto de contextos (lo que retaría, entre otras, la suposición de que en la teoría de Simmons toda oración es sólo verdadera o sólo falsa en algún contexto). Cabe la posibilidad de que R_μ *no forme* parte de los contextos sobre los que cuantifican μ y sus evaluaciones. Pero, si esto es así, hemos renunciado a la universalidad de los lenguajes coloquiales que supuestamente contienen a LO y ML, ya que estamos introduciendo restricciones en el dominio de sus cuantificadores para *evitar la expresión* (justo en el sentido denostado por Simmons, véase 1993, p. 164, n. 4) del mentiroso contextual en su forma más general, cuando habla sobre *todos* los contextos *sin excepción alguna*.

Cabe una tercera posibilidad que Simmons intenta introducir con su analogía entre μ y h. Hay oraciones que son *independientes del* *contexto* porque (1) *reciben la misma evaluación semántica en todo contexto*, como ocurre con h y también con μ, según Simmons. Esto hace que considerar el contexto sea *irrelevante* a la hora de evaluar μ y h. Esta afirmación es probablemente cierta en el caso de h porque h es independiente del contexto en un segundo sentido: (2) h *no habla* (ni directa ni indirectamente) de hechos relativos a *sus propios* contextos de proferencia sino de los contextos de proferencia de '2+2 = 4' (y '2+2 = 4' no habla ni de sus propios contextos de proferencia ni de los de h). Está claro que es posible hacer "abstracción" del contexto de proferencia en el caso de h, pero μ *no es* independiente del contexto en el sentido (2). μ *habla* de *sus* contextos de proferencia y evaluación y éstos han de tenerse en cuenta a la hora de evaluarla. Esto es así porque hay un tercer sentido en el que ni μ, ni h, ni *ninguna* otra oración son independientes del contexto, sobre todo en una teoría que versa sobre la evaluación de oraciones [*token*] (y que sólo así consigue asignar valores clásicos a oraciones [*type*] típicamente paradójicas para resolver la paradoja básica del mentiroso reforzado): (3) Ninguna oración puede ser evaluada "*fuera de todo contexto posible*", toda evaluación semántica se efectúa y se expresa *en un contexto*. El contexto de proferencia de una oración *p* (y, en su caso, los contextos de los que hable) se encuentran *involucrados* en su evaluación semántica, aunque resulte irrelevante considerarlos (como ocurre con h). Para que el contexto fuese irrelevante en el caso de μ, μ tendría que evaluarse "fuera de todo contexto" (de este modo no le afectarían ni su contexto de proferencia ni aquellos de los que habla), pero esa idea parece absurda (especialmente para una teoría de oraciones [*token*]).

Como hemos visto, interpretar 'verdadero$_{obj}$' ('V$_{obj}$') como "verdadero$_{\alpha}$ para algún contexto α" o como "independiente del contexto", presenta serios problemas de comprensión y es dudoso que resuelva el problema. En cualquier caso, podemos ignorar *qué significa* 'verdadero$_{obj}$'. En la práctica, otra posibilidad sería considerar 'verdadero$_{obj}$' ('V$_{obj}$') como un predicado *diferente* de 'verdadero' ('V'): 'V' $\in$ LO y 'V$_{obj}$' $\in$ ML. Ambos satisfacen versiones distintas del esquema tarskiano: PEV y PEV$_{obj}$. Pero aquí las dudas se multiplican de nuevo, esta vez con respecto a LO y ML.

Simmons afirma que la relación entre LO y ML no es de tipo tarskiano, ML no es un lenguaje esencialmente más rico que LO en el que éste se halle incluido. Es cierto que ML contiene predicados *no sensibles a variaciones contextuales* que no son expresables en LO ('V$_{obj}$', etc.). Desde este punto de vista, existen *criterios semánticos* para identificar oraciones de ML que no pertenecen a LO (aquellas que contienen *predicaciones* de 'V$_{obj}$' o de otros predicados similares: o_0 = 'Aquella oración es V$_{obj}$' $\in$ ML). Sin embargo, las variables de LO también pueden ser sustituidas por *oraciones de ML*. Así pues, o_1 = 'o_0 es V', *es una oración de* LO pese a contener el nombre, 'o_0', de una oración de ML. A diferencia de los lenguajes de la jerarquía tarskiana, LO tiene recursos para referir, no sólo oraciones de LO, *sino también oraciones y términos* de ML. Por otra parte, hay oraciones como μ^* = 'μ^* no es verdadera$_{obj}$' que son *singularidades* para 'V$_{obj}$', pero que *pueden* *aparecer en la extensión de* 'V' en algún contexto.[51] Desde este punto de vista, la distinción LO-ML no es tarskiana.

Si la distinción LO-ML no es tarskiana, entonces ¿por qué hemos de suponer que es *consistente*? Simmons no incluye ningún resultado formal en este sentido. Pero otro problema serio es cómo determinar a qué lenguaje pertenece cada oración. Es dudoso que esto se pueda llevar a cabo sólo sobre la base de criterios semánticos si renunciamos a jerarquías de lenguajes de complejidad semántica creciente, como Simmons. El problema se agrava cuando pensamos que LO y ML son, supuestamente, parte de lenguajes naturales como el castellano (o el inglés en términos del cual define Simmons LO y ML). ¿Cómo decidir a *qué* lenguaje pertenece una oración castellana en la que introduzcamos conectivas entre una oración de LO y otra de ML, i.e., una oración con predicados semánticos sensibles al contexto $[V([p]_c)]_{c'}$, y otra con predicados "independientes del contexto", $[V_{obj}([p]_c)]_{c'}$, como ocurre en: $[[V([p]_c)]_{c'} \wedge [V_{obj}([p]_c)]_{c'}]_{c'}$? Presumiblemente, la oración pertenece a ML, pero no está claro. Por último, aunque tuviésemos una respuesta para todas estas cuestiones, está claro que, en aras de la universalidad de los lenguajes naturales, hemos de ser capaces de expresar en castellano oraciones como: μ^* = 'μ^* es siempre (i.e., en cualquier contexto o en cualquier "evaluación no contextual") una singularidad para cualquier predicado "verdadero" (V, o V_n, o V_{obj}, ...), sensible o no al contexto,

[51] "To speak metaphorically, our ordinary context-sensitive uses of 'true' arch over the theory [ML] of that usage. But the extension of 'true_{obj}' does not: For example, it will not include any sentence that itself involves the predicate 'true_{obj}'. So, given the

presente en LO o ML'. La oración μ^* afirma de sí misma que no pertenece a la extensión de 'verdadero' bajo ninguna circunstancia contemplada por Simmons. Es, por tanto, una paradoja reforzada y sólo se puede bloquear evitando su expresión o utilizando alguna de las salidas clásicas examinadas en otras teorías (apelar a otra teoría o lenguaje y formar una jerarquía, etcétera).[52]

4.5.4 Teorías revisionistas con bucles cíclicos (Gupta y Herzerberger)

Por último, consideraremos dos propuestas similares, aunque independientes, debidas a Gupta (1982, desarrollada con posterioridad en Gupta y Belnap, 1993) y Herzerberger (1982b), aunque nos centraremos sobre todo en la propuesta de Gupta.[53] Más que una "terapia" o una solución al problema de la paradoja del mentiroso, ambas teorías están interesadas en ofrecer un "diagnóstico" adecuado o, al menos, un marco general que nos permita entender y describir el problema con precisión: cómo se producen las paradojas, cómo se comportan, etcétera. Partiendo

---

predicate 'true_{obj}' and the extension of any context-sensitive use of 'true', we cannot say that the extension of one is more comprehensive than that of the other." (p. 175)

[52] En Hardy (1997) se exponen tres contraejemplos para la teoría de Simmons: (1) La teoría evalúa de forma contraintuitiva ciertas oraciones; (2) hay oraciones, en contra de lo que afirma Simmons, que nunca llegan a ser verdaderas o falsas en ningún contexto; (3) Existen oraciones "diagonales" cuya verdad sólo se sitúa coherentemente fuera de la teoría de Simmons (i.e., la teoría no es "universal").

[53] En trabajos anteriores (1970b), Herzerberger defiende otras teorías en las que usa semánticas no clásicas. Las soluciones de Herzerberger (1982b) y Gupta (1982) son técnicamente muy similares, aunque la justificación filosófica de Gupta (especialmente en Gupta y Belnap, 1993) es mucho más rica y el alcance de su teoría, mayor. Un buen resumen de ambas teorías, de sus resultados formales y de los problemas que encaran se puede encontrar en Simmons (1993, pp. 62-9), dada la complejidad de las mismas, seguiremos aquí, en gran medida, la sencilla exposición de Simmons. También se

de un lenguaje L con un predicado 'verdadero' ('V') que satisface PEV,[54] tanto Gupta como Herzerberger recurren, al igual que Kripke, a una jerarquía de interpretaciones, I_α, que pretende imitar el modo intuitivo en que evaluamos las oraciones de nuestro lenguaje. En estas teorías, un modelo general, $M_\alpha = (M, I_\alpha)$, de L consta (1) de un modelo, M, constante pero restringido a todos los términos y oraciones que quedarían en L tras eliminar los términos 'V' y 'F' ('falso'); y (2) de un conjunto, I_α, de oraciones de L (sin restricciones) que determina la extensión de 'V' en M_α. La novedad en este caso consiste en que, contrariamente a Kripke 1975, se cumple en M_α (para todo α) que ninguna interpretación, I_α, de 'V' es parcial: toda oración, p, es verdadera ($p \in I_\alpha \leftrightarrow$ 'p' es V) o falsa (p $\notin I_\alpha \leftrightarrow$ 'p' es F) y, por tanto, las conectivas lógicas empleadas son las *clásicas*. Gupta justifica tanto la jerarquía de interpretaciones como su carácter clásico apelando a la idea de que, aprender el predicado 'V', comporta aprender un *proceso de "revisión"*.

A diferencia de lo que ocurre con otros predicados, 'P', con los que meramente asociamos procesos de "*aplicación*" (i.e., procesos que nos permiten determinar si *x es* P o *no es* P), en el caso de 'V', partimos de una hipótesis inicial con respecto a su extensión: una interpretación inicial, I_0, de 'V' que sitúa algunas oraciones de L en I_0 y el resto en I_0^- (la extensión de 'F'). A continuación, *revisamos* nuestra hipótesis de partida mediante un proceso que nos permite corregirla y mejorar nuestra

pueden encontrar resúmenes de las teorías y de los resultados formales en Visser (1989, esp. pp. 673-80), McGee (1991, pp. 127-48) y Barba (1998).

interpretación de 'V'.[55] Para dos interpretaciones cualesquiera, I_α e I_β ($\alpha < \beta$), I_β es siempre *mejor* candidato que I_α para interpretar 'V' ya que I_β está mejor situada que I_α en el proceso de revisión y, por tanto, refleja con más precisión la extensión de 'V'.

La teoría revisionista de Gupta –Herzerberger ofrece una teoría muy parecida– tiene, por supuesto, que dar cuenta de las paradojas. Todas las interpretaciones de 'V' son clásicas. Así pues, ¿qué pasa con μ = 'μ no es V'? Supongamos que $\mu \in I_0$ (μ es V como hipótesis de partida). Si esto es así, al revisar μ en I_1 concluiremos, aplicando PEV (entendido como una definición de 'V', no como un bicondicional), que $\mu \notin I_1$ (μ no es V, primera revisión), al revisar μ en I_2 concluiremos, de nuevo mediante PEV (i.e., la definición de 'V'), que $\mu \in I_\alpha$ (μ es V, segunda revisión) y así sucesivamente. En la teoría de Gupta (y en la de Herzerberger) las oraciones paradójicas similares a μ tienen una evaluación semántica *inestable* a lo largo del proceso de revisión. Las oraciones no patológicas son, en cambio, *estables*. En un momento determinado su valor de verdad se fija y permanece invariable durante todo el proceso revisionista. Para reflejar como funciona el proceso de

---

[54] En realidad, PEV no se formula en la teoría de Gupta mediante un bicondicional material. Las ejemplificaciones de PEV se consideran *definiciones* parciales del predicado 'verdadero'.

[55] "[W]e can say that underlying the use of words such as 'red' there is an application procedure that divides objects into two classes, those objects to which the word applies and those objects to which it does not apply [...] In contrast, I am suggesting that underlying our use of 'true' there is not an application procedure but a revision procedure instead. When we learn the meaning of 'true' what we learn is a rule that enables us to improve on a proposed candidate for the extension of truth. It is the existence of such a rule, I wish to argue, that explains the characteristic features of the concept of truth." (Gupta, 1982, p. 212).

revisión, tanto Gupta como Herzerberger construyen una función "salto" (idéntica a la de Kripke), S, que les permite pasar de una interpretación (o conjunto de interpretaciones) a la siguiente. La idea fundamental es la siguiente: Toda interpretación I_α *declara* como verdaderas a una serie de oraciones de M_α, así pues, la interpretación sucesora $S(I_\alpha) = I_{\alpha+1}$ declarará como verdaderas a aquellas oraciones que *son* verdaderas en M_α (corrigiendo, si es preciso, la declaración anterior, i.e., la extensión de I_α). El proceso de revisión es transfinito, así pues, ¿qué ocurre con las interpretaciones en estadios transfinitos? En el caso de Herzerberger, si λ es un ordinal límite, entonces $S(\{I_\beta: \beta < \lambda\}) = I_\lambda =$ límite inferior de I_β (para todo $\beta < \lambda$) $= \{p: \exists\beta < \lambda, \forall\gamma(\beta \leq \gamma < \lambda \rightarrow p \in I_\gamma)\}$. En otras palabras, en I_λ se declara como verdadero todo aquello que se ha declarado como verdadero de forma *estable*, a partir de una determinada interpretación de 'V', I_β, previa a I_λ.[56] Gupta ofrece la misma definición para los estadios sucesores de la jerarquía, su definición para los estadios transfinitos es parecida, pero diferente. Gupta declara verdaderas en I_λ a todas las oraciones declaradas verdaderas de forma estable antes de llegar a I_λ y a todas las oraciones verdaderas en I_0 *que no han sido declaradas falsas de forma estable* antes de llegar a I_λ, el resto son falsas y pertenecen a I_λ^-.

[56] "[A]t any transfinite limit stage what will be declared true will be exactly those statements which have been declared true at all sufficiently advanced previous stages" (Herzerberger, 1982b, p. 142).

Gupta y Herzerberger consiguen para S resultados parecidos a los "puntos fijos" que descubre Kripke en su teoría.[57] Si describimos las "oraciones verdaderas estables" (VE) como aquellas cuya verdad permanece *inalterable* a partir de I_δ (para un ordinal δ): VE = $\{p: \exists\delta\ \forall\gamma > \delta\ (p \in I_\gamma)\}$, entonces es posible mostrar que: (I) Existe un ordinal límite λ tal que I_λ = limite inferior de I_β (para todo $\beta < \lambda$) = VE. Esto es, *existe una interpretación I_λ de* 'V' (en un estadio transfinito de la jerarquía) *que declara como verdaderas sólo a las oraciones establemente verdaderas.* Obviamente, esto *no acaba con las fluctuaciones*, en el estadio $I_{\lambda+1}$ la oración del mentiroso, μ, y otras oraciones paradójicas serán declaradas verdaderas de nuevo junto a las verdades estables. Sin embargo, Gupta y Herzerberger consiguen demostrar que: (II) Dada una interpretación I_α, existe siempre un ordinal β tal que $I_{\alpha+\beta} = I_\alpha$.[58] Esto significa que, *partiendo de I_α, y tras β revisiones ($I_{\alpha+\beta}$), lo que obtenemos invariablemente es I_α de nuevo* o, en otras palabras, revisar I_α genera un *bucle* de longitud β que se vuelve a cerrar en I_α. De ese modo, *todo* el proceso de revisión *se estabiliza* y nos permite efectuar una clasificación de oraciones apelando a las diferencias de comportamiento que muestran con respecto al segmento estable del proceso de revisión: oraciones paradójicas, verdades estables, falsedades estables, etcétera.

[57] Técnicamente, es meritorio porque dos de las características que permiten a Kripke mostrar que s (su función "salto") tiene puntos fijos es que se cumple en general que para $\alpha \leq \beta$, $I_\alpha \subseteq I_\beta$ y, además, s es monótona, $I_\alpha \subseteq I_\beta \rightarrow s(I_\alpha) \subseteq s(I_\beta)$. Sin embargo, nada de esto se cumple en las teorías revisionistas, ya que las oraciones paradójicas como μ pueden entrar y salir de la extensión de 'V' en cada salto de I_α a I_β.

[58] Para ver detalles formales consúltese cualquiera de las obras citadas en la nota 53 (por su simpleza y claridad, he seguido fundamentalmente la exposición de Simmons).

Hemos dicho que el objetivo de este tipo de teorías es más "diagnóstico" que "terapéutico". Gupta ofrece una teoría de la verdad que explica cómo se comportan las paradojas y cómo afectan a nuestra concepción de lo que es verdadero. Su teoría mejora, además, la de Kripke en algunos sentidos. Ciertas oraciones que Kripke evalúa como no fundamentadas (carentes de valor de verdad) son, en cambio, intuitivamente verdaderas para Gupta (1982, p. 209-10), intuición que la teoría revisionista respeta. Ejemplos: (1) La enunciación general del principio de no contradicción para el predicado 'V': $\forall x \neg(V(x) \wedge \neg V(x))$, es una oración no fundamentada en la teoría kripkeana (ya que 'V' es ineliminable en ella), pero es una verdad estable en la teoría revisionista; (2) consideremos las siguientes oraciones pronunciadas por Juan y Pedro: (Juan dice): (A) Nada de lo que dice Pedro es verdad; (B) Algo de lo que dice Pedro es verdad. (Pedro dice): (C) Al menos una de las cosas que dice Juan es verdad. En la teoría de Gupta B y C son verdades estables, mientras que A es falsa (en la de Kripke A-C son oraciones no fundamentadas).

Ni Gupta ni Herzerberger pretenden, sin embargo, ofrecer un "antídoto" contra las paradojas, se limitan a construir una teoría que identifica las oraciones paradójicas en un modelo interpretativo estable y describe su comportamiento. "Por desgracia", ambos proyectos (el diagnóstico y el terapéutico) están más estrechamente relacionados de lo que parece. Después de todo, si fuésemos capaces de diseñar una teoría que *identificase* todas las apariciones de oraciones (proposiciones, etc.) paradójicas, podríamos diseñar también una teoría que las aislase y las "extirpase" de nuestro lenguaje para siempre, pero esa posibilidad es precisamente lo que cuestionan las paradojas reforzadas. También Gupta y Herzerberger se enfrentan a este problema. Ambos ofrecen sendas teorías que hablan sobre un lenguaje *L* en el que podemos encontrar un predicado, 'verdadero', que satisface PEV. Pero en este lenguaje no es posible encontrar el predicado '*establemente verdadero*', este predicado (al igual que 'paradójico', 'establemente falso', etc.) pertenece *exclusivamente* al metalenguaje *ML* en términos del cual hemos ofrecido una teoría para *L*. Y como es de esperar, estos predicados son objeto de sus propias paradojas: μ_{ML} = 'μ_{ML} no es establemente verdadera' (véase Simmons 1993, pp. 66-9). Estas paradojas tienen lugar en *ML*, al igual que μ ocurría en *L*, pero mientras que *ML* puede describir coherentemente el comportamiento de μ, *no puede describir coherentemente* el comportamiento de μ_{ML}. No puede hacerlo, en primer lugar, porque *ML* versa sobre *L* y 'establemente verdadero' no es un predicado de *L*, pero tampoco podría aunque el predicado problemático perteneciese a *L*. Si ampliásemos *L* para expresar *ML* (i.e., si convirtiésemos a *L* en *ML*) nos encontraríamos con el siguiente problema: nuestra teoría clasificaría μ_{ML} como "establemente verdadero" y como "no establemente verdadero" (i.e., $\mu_{ML} \in VE$ y $\mu_{ML} \notin VE$). Es cierto que podríamos construir una teoría *ML** con un predicado "*establemente verdadero**" tal que μ_{ML} *no fuese* establemente verdadero* en *ML* = *L*, sino paradójico. Pero, entonces, tendríamos una paradoja en *ML**: μ_{ML*} = 'μ_{ML*} no es establemente verdadero*'. Es obvio que el único modo de *expresar* de modo *coherente* una clasificación de las oraciones en "paradójicas", "verdaderas a veces", "establemente verdaderas", etc. es apelar a una jerarquía de lenguajes de riqueza semántica creciente. Y, una vez más, nos hallamos frente al mismo problema de siempre (véase 4.3.2), tal teoría no puede expresarse coherentemente *en su conjunto*, es decir, con respecto a todos los lenguajes que la articulan, so pena de caer de nuevo en las paradojas.[59]

4.6 Conclusiones generales

Una vez más, las conclusiones que podemos extraer de las teorías y argumentos discutidos no son, tal vez, tan "concluyentes" como alguno podría desear. Quizá haya serias objeciones a las teorías semánticas que hemos esbozado, pero parece claro que todas ellas, casi sin excepción, podrían acomodar esas objeciones (e incluso intuiciones propias a otras teorías) para mejorar su capacidad expresiva y explicativa. No cabe duda tampoco de que la mayoría de ellas aporta un valioso caudal de ideas que no deberíamos desechar sin más al tratar de explicar ciertas nociones básicas. Pero, pese a todo, también está claro que la insatisfacción de fondo sigue intacta: La inferencia de ciertas contradicciones no se puede evitar sin mutilar de alguna forma (difícilmente tolerable) los recursos

---

[59] Una vez más, las paradojas con las que tropieza la solución de Gupta no disminuyen los méritos objetivos de su teoría. Gupta & Belnap 1993 contiene uno de los estudios más completos y valiosos del fenómeno de circularidad viciosa que asociamos a nuestras reflexiones sobre la evaluación semántica de la paradoja del mentiroso (y otras afines). Por otra parte, ha permitido comparar formalmente el comportamiento de este tipo de paradojas con otro tipo de fenómenos asociados a la deliberación racional, por ejemplo decisiones que involucran la coordinación de los intereses de varios individuos, como ocurre con el dilema del prisionero, etc. (Chapuis 2000). Así mismo la teoría de la revisión ha impulsado el estudio de las aplicaciones que tienen los teoremas de punto

expresivos de nuestro lenguaje. Quizá podamos acomodar muchas críticas pero, a la vista de los indicios, quedan tantas en la recámara como respuestas podamos llegar a concebir. Es más, la mayoría de críticas (y de respuestas) parecen obedecer a un patrón común, como muestran las paradojas reforzadas y los problemas y respuestas genéricas que ocasionan.

Por otra parte, en el campo de las paradojas "semánticas" el desacuerdo y la insatisfacción son más patentes aún que en el campo de las paradojas "conjuntistas" (aunque, como dijimos, también aquí existe un grado notable de insatisfacción). La inmensa proliferación de teorías y el rechazo que causan en general (sobre todo en defensores de teorías rivales) las restricciones que nos imponen para expresar sus propias nociones semánticas de forma coherente, ponen de manifiesto que las paradojas están vivas. No son fantasmas del pasado, quizá nuestra conciencia se calme un poco cuando hablamos de lenguajes y teorías formales, pero tan pronto como tropezamos con el lenguaje natural o con nociones intuitivas no nos resignamos a aceptar restricciones expresivas, ... ni tampoco contradicciones necesarias.

Se podría decir, a modo de resumen, que, en mayor o menor medida, todas las teorías existentes son insatisfactorias en los sentidos especificados en 4.2. Todas incorporan elementos *ad hoc*, todas hacen sacrificios expresivos que no dejamos de percibir como arbitrarios, innecesarios y contraintuitivos. Es más, tenemos la sensación de que, llevadas al "límite", el mensaje de las teorías bordea incluso la

---

fijo en el diseño de teorías formales de la verdad (Visser 1989, McGee 1991, Barba incoherencia en numerosas ocasiones. Se nos promete una teoría general de las paradojas para un lenguaje *L* que sólo puede ser coherente si es dada en un lenguaje *ML* cuyos términos generan nuevas paradojas no resueltas, con lo que nos vemos forzados a construir un nuevo lenguaje *MML* que solucione las paradojas que ha creado *ML* al solucionar las paradojas de *L* y vuelta a empezar... Por otra parte, si intentamos aportar una teoría completamente general para *L* desde *L*, entonces tropezamos con serias inconsistencias. Muchas veces nos sentimos perseguidos por lo que podríamos llamar jocosamente (aunque no tanto), "el síndrome del Tractatus", subiendo y bajando escaleras "sin sentido" que al final no nos llevan a donde queríamos llegar, pero que parecen mostrarnos cosas cuya consistencia se desvanece, por otra parte, tan pronto como intentamos expresarlas.

Pero, ¿existe alguna solución a las paradojas? Hay una última respuesta que no hemos examinado aún, la de Priest, una respuesta atrevida, radical, y no exenta de interés (ni de repercusión en las revistas técnicas actualmente). Según Priest, la solución a las paradojas *consiste en aceptar que algunas contradicciones son verdaderas* (incluso *necesariamente* verdaderas). Una de las ventajas que ofrece esta respuesta es que, a diferencia del resto, propone una solución *uniforme* a todas las contradicciones de inclusión y, por tanto, puede añadir a sus méritos, y a la lista de críticas y agravios a sus adversarias, el hecho de ser la única que respeta el principio de solución única (PUS), la única que

1998, Martínez 2003 y muchos otros).

respeta el eslogan: “mismo problema, misma solución”. Veamos, pues, qué perspectivas nos ofrece esta salida.

5 El Dialetheismo de Priest

5.1 Introducción

En el capítulo anterior planteamos una cuestión: ¿Qué relación guarda el Esquema de Inclusión con la solución a las paradojas reflexivas? Vimos que la existencia de una estructura común a todas las paradojas examinadas planteaba dos ideas: (1) Desde cierto nivel de abstracción, todas las paradojas reflexivas son casos particulares de un mismo problema general de fondo. (2) Si esto es así, una solución al problema general proporcionará, con ligeros matices y adaptaciones, una solución *uniforme* a cada una de las paradojas que nos ocupan. (En un eslogan: mismo tipo de problema, mismo tipo de solución.) La tesis (2) fue bautizada con el nombre de Principio de Solución Uniforme (PSU). Antes de examinar el significado y la validez real de PSU, nos preguntamos qué requisitos ha de satisfacer una teoría que ofrezca una solución satisfactoria para las paradojas reflexivas. Señalamos cinco: (a) Debe ofrecer un "diagnóstico" acertado del problema y una solución que se adecue al diagnóstico (no una solución cualquiera, arbitraria y *ad hoc*); (b) debe ser coherente (consistente, libre de contradicciones); (c) debe conservar la verdad de todos aquellos enunciados que no tenemos razones para rechazar y sí para mantener (i.e., no debe mutilar nuestra visión del mundo); (d) debe dar cuenta de sí misma (evitando regresos al infinito con respecto a su propia justificación). Y, por último, y con carácter provisional, aceptamos también que (e) debería respetar PSU.

Vimos que, con independencia del diagnóstico efectuado, la mayoría de teorías optaba por utilizar cada contradicción de inclusión como una *reductio ad absurdum* de la existencia de Ω (la totalidad de los conjuntos, de las verdades, de los predicados heterológicos, de los ordinales, etc.). Vimos también que casi siempre recurrían a estrategias "paramétricas" en las que la paradoja se resolvía descubriendo cierta ambigüedad en el seno de Ω que podemos disolver mediante una jerarquía de Ωs "parametrizadas": Ω_0, Ω_1, Ω_2, ... De este modo, la contradicción desaparecía convirtiéndose en $\delta(\Omega_n) \notin \Omega_n$ y $\delta(\Omega_n) \in \delta(\Omega_m)$ donde $m \neq n$. Vimos, además, que, frente a esta estrategia, siempre era posible construir la unión de todos los Ω_β ($\beta < \alpha$) parametrizados: $\boldsymbol{\Omega^*} = \bigcup_{\beta < \alpha} \Omega_\beta$ y formular una nueva paradoja a partir de $\delta(\boldsymbol{\Omega^*})$. La única manera de evitar esta maniobra era atentando contra uno o varios de los requisitos (a)-(d), esto es, introduciendo restricciones *ad hoc* que recortan el poder expresivo de los lenguajes naturales, impidiendo la expresión de una teoría general sobre las paradojas en un lenguaje consistente, etcétera. Entre las soluciones a las paradojas conjuntistas examinamos el sistema ZF y la distinción *clase/clase propia* de von Neumann. Entre las soluciones a las paradojas semánticas, examinamos cuatro grupos: Teorías jerárquico-sintácticas (Russell y Tarski), teorías no bivalentes (van Fraassen, Martin, Kripke, McGee,), teorías contextualistas (Burge, Barwise y Etchemendy y Simmons) y teorías revisionistas (Gupta y Herzerberger). Todas ellas incorporaban ideas "paramétricas" y tropezaban con el mismo problema, elegir entre la coherencia y la generalidad expresiva. Dado que renunciar a la generalidad significaba, entre otras cosas, renunciar a la expresibilidad y a la justificación de la propia teoría dentro de sí misma, los problemas de coherencia y expresibilidad acababan entrelazándose. Por último, la evidencia de que las llamadas paradojas "conjuntistas" y "semánticas" recibían un trato diferente añadía una objeción más a su viabilidad: las teorías mencionadas no respetan el Principio de Solución Uniforme.

El objetivo de este capítulo es examinar una última teoría, el dialetheismo de Priest, según la cual ciertas contradicciones (las paradojas de inclusión) son verdaderas. Aunque esta solución parece respetar PSU, está por ver si es satisfactoria en los sentidos (a)-(d).

5.2 El dialetheismo de Priest frente a otras teorías: Coherencia, generalidad y PSU[1]

Priest nos remite a la historia de la filosofía en varios momentos a lo largo de *Beyond the Limits of Thought*[2] para hacer hincapié en una idea: Llevadas al límite, nuestras reflexiones sobre algunos de los conceptos básicos que articulan el pensamiento humano han producido siempre (y continúan produciendo) paradojas. Kant y Hegel (*IC*, pp. 3-4; *BLT*, pp. 73-109) fueron quizá los primeros filósofos que cobraron conciencia de este fenómeno otorgándole un papel central en sus respectivas obras. Kant habló de "antinomias de la razón pura" y Hegel de contradicciones "en el mundo" y de "lógica dialéctica". La resistencia

[1] Aunque el dialetheismo no sea muy popular, cabe decir que Priest no está solo. Armour-Garb y Beall (2002, p. 217), por ejemplo, se declaran dialetheistas.

a desaparecer que ciertas paradojas muestran después de tantos años e intentos fallidos por erradicarlas sugiere que algunas de las contradicciones que dichas paradojas establecen son *verdaderas*. Priest bautiza con el nombre de "*dialetheias*"[3] a este tipo de contradicciones verdaderas y defiende su existencia. Al defenderlas, no apela a Kant o Hegel, ya que, según Priest, la mayoría de antinomias kantianas constituyen falsas aporías (*BLT*, pp. 88-99) y algunos de los conceptos empleados por Hegel para ilustrar la existencia de contradicciones en el mundo (el de infinito, por ejemplo) no tenían en su época el contenido claro que les atribuimos hoy en día (*BLT*, p. 109). En su defensa de las dialetheias, Priest se centra sobre todo en conceptos básicos de nuestro lenguaje, en expresiones tales como '... es verdadero', 'No es el caso que ...' o '... pertenece a ...' fundamentales para el desarrollo de un sinfín de prácticas teóricas (desde las más básicas hasta las más complejas). Habitualmente asociamos con estas expresiones una serie de principios que consideramos esenciales para su comprensión y posterior uso (principios analíticos y constitutivos de su significado), pero que, no obstante, generan paradojas irresolubles (*IC*, pp. 4-5). Entre esos principios, Priest incluye el Esquema-V tarskiano, el Principio de Abstracción (PEV y PEA respectivamente en 4.3.1) y lo que podríamos llamar "Principio Esquemático de la Negación" (PEN):

---

[2] Siguiendo el mismo criterio que en capítulos anteriores, utilizaré las abreviaturas *IC* y *BLT* para citar dos obras de Priest: *In Contradiction* (1987) y *Beyond the Limits of Thought* (2002) respectivamente.

(PEN) '$\neg\alpha$' es verdadera $\leftrightarrow$ 'α' es falsa.

Cuando combinamos estos principios incuestionables con otros principios básicos de inferencia y con una serie de recursos expresivos inherentes a los lenguajes naturales (y a otros no tan generales) obtenemos "demostraciones" intuitivas de contradicciones en diferentes ámbitos. Si queremos preservar la validez de nuestros razonamientos en estas áreas y, por consiguiente, la verdad de las premisas en la conclusión, tenemos dos opciones: O bien rechazamos la validez de alguno de los principios (o la verdad de alguno de los supuestos) involucrados en el razonamiento, o bien aceptamos la verdad de la conclusión que nos impone: una contradicción. Priest opta por lo segundo.

Obviamente, defender la existencia de contradicciones verdaderas (¡demostrables!) va en contra de todo lo que uno aprende en cualquier curso de lógica y también en contra del magisterio (según Priest, no seriamente cuestionado hasta ahora) de Aristóteles en el libro Γ de la *Metafísica*, donde defiende el principio de no contradicción.[4] Así pues, Priest necesita convencernos de muchas cosas antes de que aceptemos su propuesta. Pero Priest no ignora este hecho y señala cuáles son los principales frentes en los que debe combatir el dialetheismo. Su teoría debe (*IC*, pp. 5-6): (1) Ofrecer indicios a favor de la existencia de

---

[3] "[A] dialetheia is any true statement of the form: α and it is not the case that α." (*IC*, p. 4)

[4] Priest (1998a) examina detalladamente (y rechaza) los argumentos ofrecidos por Aristóteles a favor del principio de no contradicción en la *Metafísica.*

dialetheias y de la validez (pese a su inconsistencia) de los principios y supuestos que las establecen. (2) Ofrecer una explicación alternativa de la noción de coherencia, desligándola de la noción de consistencia. Si el Principio de No Contradicción (PNC) *no* es universalmente válido, ser incoherente no es lo mismo que ser inconsistente. (3) Para mostrar que coherencia y consistencia se pueden separar, hay que mostrar en este contexto que ser inconsistente *no implica* ser incoherente y, sobre todo, que de una contradicción *no se sigue cualquier cosa*. El principio lógico de "*Ex Contradictione Quodlibet*" (ECQ) o "*Explosión*", como Priest lo llama a veces (1998b, p. 411), *no es válido*.

A cambio del *tour de force* que nos promete, Priest sólo pide una concesión: que no incurramos en una *petición de principio*, esto es, que no esgrimamos o aceptemos como una objeción válida contra el Dialetheismo la tesis de que no puede haber contradicciones verdaderas *sin ofrecer un argumento independiente* a favor de dicha tesis (i.e., de PNC).[5]

[5] "It might be suggested that no further investigations are necessary, since the contradictions produced show *ipso facto* that the principles [PEV, PEA, PEN] are not correct. Not so. The inconsistency of our linguistic principles is the very thesis I am affirming. Hence, in discussions of what these principles are, consistency cannot be invoked as a regulative principle without begging the question against me." (*IC*, p. 5). Ésta es una falta de la que, a menudo con razón, son acusados quienes objetan a Priest. La idea es recurrente (*BLT*, p. 276). No sólo sería ilícito presuponer la validez de PNC, sino también la de ECQ ("explosión"): "[T]he failure of explosion is a plausible logico-metaphysical view, and one cannot simply assume otherwise without begging the question." (Priest 1998b, p. 416; la misma idea se repite en Priest 1994b, p. 341).

5.2.1 Indicios o pruebas a favor del Dialetheismo de Priest

La primera estrategia a seguir por Priest es mostrar el fracaso histórico de todas aquellas teorías que, a lo largo de siglos (sobre todo en el XX y, con particular intensidad, en su último tercio) han intentado inútilmente aportar una solución consistente y satisfactoria a las paradojas semánticas y conjuntistas. Si bien su fracaso no establece por sí mismo la verdad del dialetheismo, sí ofrece indicios *empíricos* nada despreciables a favor de la posible existencia de contradicciones verdaderas.[6] En el capítulo anterior nos dedicamos a mostrar en qué sentido las soluciones existentes podían considerarse insatisfactorias con respecto a una serie de *desiderata* resumidos en 5.1. Todas las teorías allá examinadas se enfrentaban a un dilema: O restringían el ámbito y la capacidad expresiva y explicativa de sus teorías de tal manera que algunos de los conceptos empleados por la propia teoría quedasen fuera de su alcance (i.e. renunciaban a la reflexividad); o se aferraban a la generalidad o universalidad expresiva y explicativa de sus teorías (y, por tanto, a la reflexividad) y se convertían así en víctimas indefensas de nuevas paradojas reforzadas. La primera opción es insatisfactoria porque no consigue explicar el fenómeno de las paradojas en general (sobre todo en los lenguajes naturales) y porque no justifica la coherencia de los conceptos en los que se apoya la teoría relevante. La segunda opción es igualmente insatisfactoria porque todas las teorías examinadas prometían

---

[6] "If my argument for accepting the paradoxes were merely that no one had yet solve them, then my position, though plausible, would not be very tempting." (Priest 1979, II.1, p.220). A continuación añade: "However, I believe that there are theoretical

erradicar las contradicciones y, sin embargo, nos permiten inferir nuevas.[7] Los dialetheistas optan por el segundo cuerno del dilema, pero, a diferencia de las teorías apuntadas en 4, la inferencia de ciertas contradicciones no les supone ningún problema, ya que, para comenzar, nunca pretendieron deshacerse de ellas sino justificar su verdad.

Al margen de este problema recurrente existen otros. Las teorías alternativas al dialetheismo suelen imponer más tarde o más temprano algún tipo de restricción *ad hoc* –i.e., justificada *únicamente* a partir de su eficacia para bloquear la inferencia de una contradicción– a la validez intuitiva y aparentemente general de principios fundamentales como PEA, PEV y PEN. Los dialetheistas, en cambio, no recurren a restricciones *ad hoc* y contraintuitivas de estos principios porque no

---

reasons why paradoxes must be accepted; must, that is, if we are to get to grips with the notion of mathematical provability (in the naïve sense)." (Véase 5.2.1.1)

[7] Una de las exposiciones más claras de este dilema se encuentra en Priest 1999 (pp. 109-10). Priest repite esta idea continuamente, especialmente en *IC*, pp. 29-31, 47-8. "For each of the 'solutions' we have considered, it transpires that the solution generates concepts which allow a different version of the paradox to be given. Thus nothing has been gained. Ultimately the only consistency-generating move is to deny that those concepts in which the solution is expressed (which is English) are not expressible in the language for which the solution is being given, which is just an admission that the problem of showing English to be consistent has not been shown." (*IC*, p. 29) Más adelante, añade: "The claim that consistency ultimately drives one to self defeating metalanguage can be supported by other reasons too. The paradox solving problem is to produce a consistent theory that can express its own semantic notions. But this is a classical chimera: If a theory is to give an account of its own semantics, it must give an interpretation of some kind for the language of the theory. Then, to show that it is a semantics for the theory with respect to the interpretation. But (classically) soundness implies consistency, and, provided that the language is sufficiently strong, we know that an internal consistency proof is impossible by Gödel's second incompletness theorem. What the Gödel incompletness theorem thus shows is that, classically, consistency can be maintained only by giving the semantics of a theory in a different theory." (*IC*, p. 30). Con respecto a las paradojas conjuntistas se expresa en términos parecidos (*IC*, pp. 47-8).

encuentran ningún problema en aceptar su validez general y las inconsistencias que de ella se siguen.[8] Por último, las paradojas que surgen de estos principios comparten una misma estructura, la que refleja el Esquema de Inclusión (hemos apoyado esta tesis en los capítulos 2 y 3), y cabe suponer, por ello, que son en realidad diferentes versiones de un mismo problema general de fondo. Todas las teorías examinadas ofrecen soluciones dispares a las paradojas de inclusión, sin embargo, el dialetheismo de Priest ofrece una solución uniforme y sencilla para todas ellas: Allá donde se produzcan, las contradicciones que establecen son verdaderas. Así pues, sólo el dialetheismo respeta el Principio de Solución Uniforme (PSU).[9]

Obviamente, no reproduciremos aquí la discusión de los capítulos precedentes, pero sí es pertinente añadir lo que Priest considera una "prueba" más de la corrección de su teoría. Se trata de una "paradoja" que no hemos abordado hasta ahora (básicamente porque pocos la

[8] "A paradox is an argument with premises which appear to be true and steps which appear to be valid, which nevertheless ends in a conclusion which is false. A solution would tell us which premise is false or which step invalid; but moreover it would give us an *independent reason* for believing the premise or the step to be wrong. If we have no reason for rejecting the premise or the step other than that it blocks the conclusion, then the 'solution' is *ad hoc* and unilluminating. Virtually all solutions to the paradox fail this test and this is why I say that no solution has yet been found." (Priest 1979, I.2, p. 220 –cursiva en el original). Priest aún mantiene este punto de vista.

[9] "[...] Russell's instinct was right. All the traditional paradoxes of self-reference are inclosure contradictions. That is, the structure described in the Inclosure Schema explains all of these contradictions. Hence, by the PUS [Principle of Uniform Solution], one should be satisfied with nothing less than a unified solution to the family." (*BLT*, p. 167). Más adelante añade: "[N]o single post-Ramsey solution applies to all the inclosure contradictions. Even if the orthodox solutions worked where they were designed to [...], this observation would be sufficient, on its own, to sink them." (Ibíd., 167).

consideran como tal) y que viene a añadir nuevos argumentos a favor de la posibilidad de *inferir contradicciones verdaderas*.

5.2.1.1 El teorema de Gödel: Teorías inconsistentes de la demostración

Según Priest, el primer teorema de incompletud de Gödel esconde en el fondo una paradoja (una contradicción de inclusión)[10] que afecta a nuestra noción intuitiva o "naif" de demostración. Tal y como la entienden los matemáticos, la noción de "prueba" o "demostración" nos remite a una serie de procedimientos informales por medio de los cuales establecemos la verdad de ciertos enunciados matemáticos a partir de otros cuya verdad conocemos previamente. Esto es, demostramos *p* apelando a una serie de reglas de inferencia y enunciados *q*, *r*, *s*, ... cuya verdad nos es dada. Para evitar regresos al infinito, contemplamos la existencia de un tipo de enunciados no demostrables cuya verdad se toma como "básica" a la hora de fundamentar la demostración de otros enunciados matemáticos (*IC*, p. 50).[11] Según Priest (*IC*, p. 51; 1979, p. 221), una teoría adecuada de nuestra noción intuitiva de "demostración" debe cumplir las condiciones del primer teorema de incompletud de Gödel, lo que, como veremos, conlleva ciertos problemas.

---

[10] Véase *IC*, capítulo 3 (especialmente, pp. 58-60) o Priest 1979 (pp. 220-5). Esta tesis se apoya también en Priest (1994b, pp. 43-4). En *BLT*, p. 144 (y Priest 1998c, p. 836) Priest defiende, además, que la "paradoja" de Gödel encaja en el Esquema de Inclusión. Aunque nunca muestra exactamente cómo, sí da algunas pistas al indicar que es una variante de la paradoja del conocedor [16].

[11] En Priest 1979 (p. 221), estos enunciados reciben el nombre de axiomas y son verdades "autoevidentes".

Teorema de Gödel (TG): Si la teoría T puede representar todas las funciones recursivas, y su relación de "demostrabilidad" es recursiva, entonces existe una fórmula φ tal que: (i) Si T es consistente, φ no es demostrable en T (y ¬φ tampoco); y (ii) si los axiomas y reglas de T son intuitivamente correctos [*sound*], podemos establecer un argumento intuitivamente correcto [*sound*] en virtud del cual φ es verdadera.[12]

Para mostrar que nuestra teoría intuitiva de la demostración satisface las condiciones que permiten aplicar TG, Priest defiende dos tesis:

(1) *Podemos formalizar nuestra teoría "naif"* (T) *de la demostración.* Aunque las demostraciones matemáticas (y, en particular, las de la aritmética) se den habitualmente en un lenguaje informal, esto obedece a criterios prácticos. Según Priest (*IC*, p. 51; 1979, p. 221), el fragmento del castellano en el que expresamos nuestra teoría intuitiva de la demostración para la aritmética podría convertirse en una teoría formal, T, que codifique nuestros teoremas, enunciados básicos, fórmulas, funciones, etc. mediante números naturales. En particular, T podría *representar y codificar cualquier función recursiva.*

(2) *La teoría,* T, *puede representar la relación* "*x* es una *demostración* de *y* en T" *que, además, es recursiva.* Según Priest (*IC*, pp. 51-2), algo es una "prueba" sólo si podemos *reconocerla* como tal mediante algún tipo de test, en caso contrario, dada una prueba real, no sabríamos cómo establece lo que de hecho establece. Quizá no sepamos

[12] Ésta es la versión del teorema que facilita Priest en *IC*, p. 49.

si existe una prueba (o una refutación) de un enunciado *p*, pero, si nos enfrentamos a algo, un *x*, que resulta ser una prueba de *p*, debemos ser capaces de *decidir* (mediante algún tipo de procedimiento algorítmico) que *x* es, efectivamente, una prueba de *p*. Es más, si "demostrar" no fuese una práctica descriptible de forma recursiva, no podríamos aprender *qué* es una demostración, ni, por ende, a demostrar.

Ahora bien, si la teoría T que capta nuestros procedimientos intuitivos de demostración satisface (1) y (2), entonces podemos aplicar TG sobre ella obteniendo resultados paradójicos (*IC*, p. 56; Priest 1979, p. 221): Si T es consistente, entonces, por (i), existe una fórmula φ que *no es demostrable en* T (¬φ *tampoco*). Pero, por (ii), *podemos demostrar intuitivamente* que φ *es verdadera*. Priest se desmarca de las interpretaciones tradicionales de TG en dos puntos. *Primero*: Priest (*IC*, p. 59; 1979, pp. 223-4) afirma que la verdad de φ se puede demostrar *en* T porque T es semánticamente cerrada (contiene el esquema tarskiano, PEV, para su propio predicado de verdad y puede demostrar 'φ' a partir de 'φ es verdadera'). *Segundo*: Priest argumenta en (2) que T, nuestra teoría *intuitiva* de la demostración, puede ser *mecanizada* (i.e., formalizada y descrita recursivamente). La "paradoja de Gödel" se sigue de estas afirmaciones: Si T es consistente, φ no se puede demostrar en T por (i), sin embargo, por (ii), φ se puede demostrar en T. Así pues, T debe ser inconsistente.

Contra la primera afirmación, las interpretaciones consistentes de TG afirman que *ni* la demostración intuitiva de φ *ni* algunos de los principios que la facilitan (por ejemplo, PEV) *pueden pertenecer a* T o a otra teoría axiomatizable que satisfaga los requisitos de TG. Por consiguiente, contra la segunda afirmación, si nuestra teoría intuitiva permite inferir φ y es consistente, dicha teoría no es mecanizable. Pero veamos con más detenimiento cómo se llega a una contradicción en T examinando un esbozo de la demostración de TG. Tomemos el siguiente predicado de T, cuyo código es *s*:[13]

(*s*) $\neg\exists x\, Demo(x, y)$

Demo(*x*, *y*) representa la relación de demostrabilidad en T (*x* se sustituye por el código de una demostración en T de la oración *y*). Como la relación de demostrabilidad es recursiva, *s* se puede representar en T sin problemas, siendo su interpretación natural: "*y* no es demostrable en T". Consideremos ahora la oración, (con código) *g*, que dice *de sí misma* que no es demostrable en T: g = '$\neg\exists x\, Demo(x, g)$'.[14] Si *g* fuese demostrable, existiría una demostración, (con código) *n*, de *g* en T y se cumpliría: T $\vdash$ *Demo*(*n*, *g*) ('$\vdash$' representa la relación de inferencia en T). Ahora bien, de aquí se sigue una contradicción: T $\vdash \exists x\, Demo(x, g) \wedge \neg\exists x\, Demo(x, g)$, ya que (I) $Demo(n, g) \vdash \neg\exists x\, Demo(x, g)$ y (II) $Demo(n, g) \vdash \exists x\, Demo(x,$

[13] Como es habitual, interpretamos esta fórmula abierta con una variable libre como un predicado.

[14] Para justificar la existencia en T de este tipo de oraciones autorreferentes, Priest apela a lo que denomina "*Diagonal Lemma*" (*IC*, pp.61-2) según el cual: Para toda fórmula con una variable libre φ(*x*), existe una oración γ tal que $\gamma \leftrightarrow \varphi(\underline{\gamma})$, donde $\underline{\gamma}$ es el código de la oración γ en T.

g).[15] Así pues, si g fuese demostrable, T sería inconsistente, lo que, por contraposición, nos permite establecer (i) en TG: Si T es consistente, g no es demostrable.

Para demostrar (ii) en TG debemos probar dos cosas: (A) que T es intuitivamente correcta [*sound*] y (B) que, si (A) es el caso, entonces g es verdadera. (A) se demuestra por inducción sobre la estructura de las reglas de inferencia, R, que permiten establecer $\Sigma \vdash \varphi$ en T. Comprobamos en cada caso ($R_0, R_1, ..., R_n$) que, si las fórmulas de Σ son verdaderas, también lo es la fórmula φ inferida mediante $R_{i \leq n}$ (*IC*, pp. 62-3). Una vez probado (A), establecemos (B) apelando al siguiente argumento intuitivo: Supongamos que T es consistente, según TG(i), g no es demostrable, esto es, para toda demostración n, se cumple $\neg Demo(n, g)$, de donde inferimos por generalización universal $\forall x \neg Demo(x, g)$. Ahora bien, $\forall x \neg Demo(x, g)$ es equivalente a $\neg\exists x\ Demo(x, g)$, es decir, equivalente a g. Por tanto, podemos establecer que *g es verdadera* aplicando PEV (el esquema tarskiano) sobre $\neg\exists x\ Demo(x, g)$. De este modo completamos la demostración de (ii) y, por tanto, la de TG.

Frente a TG caben dos posturas, la tradicional y el dialetheismo. Según la postura tradicional, la inferencia de una contradicción es siempre intolerable, así pues, al percatarnos en (i) de que, si g es demostrable en T, intuitivamente también lo es su negación, rechazamos

[15] (II) se sigue del principio de generalización existencial. (I) se justifica apelando al principio: (*) $Demo(n, x) \rightarrow p$ (x es un nombre de la oración p) cuyo significado intuitivo es "Si 'p' es demostrable, entonces p". Si, como afirma (2), la relación '*Demo*' es *representable* en T, (*) ha de ser válido en T, ya que decir que 'p' *es demostrable en* T significa, intuitivamente, decir que *p se obtiene en* T.

la demostrabilidad de *g*. Sin embargo, (ii) nos dice que, intuitivamente, *g* *es verdadera* y, si esto es así, podríamos *inferir* paradójicamente *g* apelando al esquema tarskiano. Para evitar este resultado la postura tradicional se ve forzada a afirmar que ni el razonamiento que establece la verdad de *g* en (ii) ni alguno de los principios que involucra (por ejemplo, PEV) pertenecen a la teoría T (i.e., ni son axiomas de T, ni son demostrables en T, ni son representables en T). De este modo, T es una teoría *consistente* pero *incompleta*, ya que no todas sus verdades son demostrables. Según el dialetheismo de Priest, por contra, todos los principios y razonamientos que aparecen en (i) y (ii) son intuitivamente correctos. Dado que T es nuestra teoría *intuitiva* de la demostración en aritmética, resulta intolerable excluir de ella estos principios (incluido PEV) y razonamientos. Sólo un argumento *ad hoc* destinado a evitar la verdad de algunas contradicciones (*g* y no *g*) podría excluirlos de T. Puesto que el dialetheismo defiende que algunas contradicciones son verdaderas (en particular, *g* y no *g*), incurriría en una petición de principio quien rechazase esta posibilidad sin ofrecer argumentos independientes a favor de la falsedad de toda contradicción.[16] Así pues, T es *inconsistente* y *completa*, todas sus verdades (incluida *g* y no *g*) son demostrables.

Sólo si interpretamos TG como Priest, podemos evitar la paradoja de afirmar que *g* es indemostrable en T y de ofrecer a continuación una demostración de su verdad en una "teoría" que incluye PEV y que, por

[16] En 5.3.7 veremos algunos argumentos independientes y objeciones a las tesis dialetheistas respecto a T.

razones *ad hoc*, no puede identificarse con T. Si Priest tiene razón, entonces TG apoya la tesis dialetheista al mostrar que nuestra teoría intuitiva de la demostración es esencialmente inconsistente y, por ende, que es razonable aceptar la verdad de ciertas contradicciones en la medida en que son demostrables.[17]

5.2.2 ¿Es coherente el dialetheismo y racional aceptar una contradicción?

El hecho de que las aproximaciones consistentes a las paradojas de inclusión y a TG sean en cierto modo insatisfactorias (por generar inconsistencias, pese a rechazar toda contradicción) sólo podría contar como una razón a favor del dialetheismo si juzgamos que es coherente y racionalmente satisfactorio aceptar contradicciones verdaderas. Recordemos que éste era uno de los principales requisitos que debía cumplir en 4.2 toda solución a las paradojas. ¿Es coherente aceptar la verdad de una contradicción? Tradicionalmente se ha asociado la noción de coherencia a la de consistencia, siendo la segunda, al menos, una condición *necesaria* para la primera: *Sólo* si *no* se contradicen entre sí, pueden nuestras creencias ser coherentes. El dialetheista acepta contradicciones, por tanto, es incoherente.

---

[17] Chihara (1984b, p. 118) advierte con razón que T no precisa ser semánticamente cerrada (o apelar a PEV) para ser inconsistente, basta con que incluya el razonamiento que demuestra $\neg\exists x\, Demo(x, g)$.

Como es natural, Priest se opone a esta idea y sugiere que hay mejores maneras de entender la coherencia.[18] Ciertamente, la consistencia interna de una teoría debería ser un criterio relevante a la hora de determinar su coherencia, pero no debería ser ni un criterio suficiente ni un criterio necesario. Lo primero se pone de manifiesto cuando reflexionamos sobre el vínculo existente entre coherencia, verdad y justificación racional. Cuando le pedimos coherencia a una teoría, le pedimos algo más que un conjunto de afirmaciones que no se contradigan *entre sí*, le pedimos afirmaciones que no contradigan *nuestra experiencia* ni los *principios* que rigen *nuestro razonamiento*. Obviamente, si esos principios incluyen el de no contradicción, entonces pedimos también consistencia, pero la consistencia no puede ser una condición *suficiente* para la coherencia. Hay conjuntos de creencias falsas que son, no obstante, consistentes. Si redujésemos "coherencia" a "consistencia", una teoría "coherente" podría estar completamente divorciada de la verdad, lo que no parece intuitivamente aceptable. Pedirle coherencia a una teoría es, pues, pedirle que satisfaga una serie de requisitos de carácter cognitivo que justifiquen nuestra creencia en la verdad de sus afirmaciones. La consistencia es, sin duda, uno de esos requisitos, pero hay otros.[19]

---

[18] Las tesis de Priest (que ahora trataremos) sobre la noción de coherencia y su relación con la noción de consistencia aparecen con claridad en Priest 2000a, pp. 312-4, aunque podemos encontrar observaciones similares a lo largo del capítulo 7 de *IC* (sobre todo pp. 119-28) y en Priest 1998b, pp. 419-24.

[19] Hablando sobre la teoría de la verdad como coherencia, Priest dice "those who endorsed the theory have held that it makes no sense to define truth in terms of some objective reality independent of our cognitive functioning: there is no such thing, or if there is, we have no access to it. If we are to have any meaningful notion of truth, this

Ahora bien, ¿es una teoría coherente necesariamente consistente? Si el principio de no contradicción (PNC) tiene validez general, entonces sí; en caso contrario, no. Así pues, la respuesta presupone, una vez más, nuestra valoración de PNC. El dialetheista niega su validez apelando, básicamente, a dos argumentos. El primero descansa (como vimos en 5.2.1 y en el capítulo 4) en la existencia de indicios empíricos a favor de la "verdad" de ciertas paradojas. Hasta ahora, ninguna teoría ha conseguido resolver por completo y de forma consistente las paradojas de inclusión que contiene el lenguaje ordinario. El segundo argumento atañe a las consecuencias que *no se siguen* de aceptar una contradicción, pero, antes de abordarlo (en 5.2.2.1), veamos qué ocurriría si el dialetheista tuviese razón. Si éste fuese el caso, la coherencia de nuestras creencias residiría en la satisfacción equilibrada de una serie de criterios o indicadores epistémicos de justificación racional. La consistencia sería uno de los criterios principales, pero, en ciertas circunstancias –muy extraordinarias, aquellas en las que nuestra experiencia nos impone recalcitrantemente la verdad de una contradicción–, tendrían más peso otros criterios. Por ejemplo, el número de principios cuya validez general podemos preservar rechazando PNC, la sencillez de nuestra solución, la concordancia con nuestra experiencia, la posibilidad de explicar otros fenómenos, etcétera. Una teoría sería coherente en la medida en que

can be defined only in terms of what we are justified in believing (maybe in the ideal limit). The criteria of coherence are therefore the criteria of justification." (Priest, 2000b, p. 313) Y añade: "Now an important part of justification of any overall theory is empirical adequacy, that is, consonance with observation" (ídem), aunque también advierte que no es el único criterio, ya que nuestras observaciones empíricas son falibles y otros criterios coherentistas intervienen en nuestra aceptación de una observación.

consiguiese *maximizar* todos estos criterios epistémicos que, por sí solos (sin tener en cuenta su relación con otros y con nuestra experiencia), no ofrecerían condiciones necesarias ni suficientes para evaluar la coherencia global de una teoría. Esta visión de la coherencia es, sin duda, atractiva y la razón de ello radica en que la coherencia es una cualidad de carácter "*holista*". Una creencia es coherente con respecto al grupo de creencias al que pertenece, no sólo a una o dos de ellas, y, aunque nuestros principios tienen un estatus especial, no dejan de ser creencias cuya validez hemos de revisar a la luz del resto de nuestras creencias.[20] La consistencia de una teoría se puede perder en el momento en que encontremos en su seno una oración, φ, y su negación, $\neg\varphi$; la coherencia, en cambio, parece más difícil de construir, pero también de destruir. Una teoría consistente podría no ser coherente si reta sistemáticamente a

[20] Esta noción de coherencia no está tan alejada de cierta visión "holista" de la ciencia según la cual la validez de una teoría reside, más que en su consistencia interna, en su coherencia *con nuestra experiencia*. Considérese la imagen quineana de la ciencia en "Two Dogmas": "The totality of our so-called knowledge or beliefs [...] is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements." (Quine, 1980b, p. 42) Quine llega a afirmar: "Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system." (Ibíd., p. 43) E insiste a continuación en que "no statement is immune to revision. Revision even of the logical law of excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle?" (Ibid., p. 43). Después de todo, Priest, nos propone, basándose en *nuestra experiencia* sobre las paradojas y para *ajustarse coherentemente a ella*, que revisemos la validez general del principio de no contradicción. (No obstante, no deja de ser cierto que Quine moderaría con posterioridad esta visión de la ciencia en lo tocante a la lógica. Esto guarda relación con los sacrificios de principios lógicos que impone la renuncia a PNC, como veremos más adelante.)

nuestra experiencia y a la validez de un elevado número de principios; sin embargo, si tiene razón el dialetheista, una teoría coherente podría soportar alguna contradicción, siempre y cuando éste fuese el único modo de preservar la validez global de una serie de principios y creencias cuyo rechazo no podemos justificar.

5.2.2.1 Principios de No Contradicción (PNC) y Ex Contradictione Quodlibet (ECQ)

Priest afirma que, dada nuestra experiencia con las paradojas, es posible, e incluso razonable y coherente (en el sentido dialetheista), aceptar *alguna* contradicción, sin embargo, también afirma que *no* es posible (ni razonable, ni coherente) aceptar la verdad de *cualquier* contradicción, ni mucho menos la de *todas*. Priest se esfuerza por mostrar que la tesis dialetheista según la cual *algunas* contradicciones son verdaderas puede y debe separarse de la tesis que afirma la verdad de *todas* las contradicciones, ambas son mutuamente independientes.[21] Si el dialetheismo defendiese la verdad de *todas* las contradicciones, tendría serios problemas. Presumiblemente, toda oración (enunciado o proposición), p, puede ser negada, $\neg p$. Así pues, afirmar que *todas* las contradicciones, p y $\neg p$, son verdaderas equivale, en la práctica, a afirmar que *todas* las oraciones (enunciados o proposiciones) son

[21] Esta idea es crucial: "What is wrong with believing *some* contradictions? I emphasize the 'some'; the question 'What is wrong with believing *all* contradictions' is quite different", and, I am sure, has a quite different answer. [...] It is important to emphasize this distinction [...], since the illicit slide between 'some' and 'all' is endemic in discussions of the question." Priest 1998b, p. 410 (Priest ilustra la idea más adelante con ejemplos (pp. 417, 418, 422, 425), véase también *IC*, p. 129; o Priest 2000b, p. 189).

verdaderas, ya que todas tendrán la forma de una afirmación, *p*, o de una negación, $\neg\, p$. Ahora bien, si este fuera el caso cualquier práctica encaminada a determinar la verdad de una oración sería inútil. Al ser todo verdadero, no tendría sentido preguntarse de algo si lo es.

Las consecuencias de este hecho serían devastadoras dado el estrecho vínculo existente entre las nociones de "verdad" y "significado". Si identificamos, por ejemplo, "significado" con "condiciones de verdad", perdería todo sentido hablar del significado de una oración, ya que sería imposible distinguir las condiciones que hacen verdadera a una oración de las que hacen verdadera a otra, todo sería verdadero bajo las mismas condiciones y, por tanto, el significado de una oración sería intercambiable con el de cualquier otra. Tampoco irían mucho mejor las cosas si definiésemos significado en términos de uso. Deberíamos entonces diferenciar las circunstancias en las que usamos una oración, o aplicamos un predicado, de las circunstancias en las que no lo hacemos, pero, como todo sería verdadero, *se cumpliría* que "esas circunstancias son las mismas para toda oración o predicado en cualquier caso". También nuestras acciones, decisiones, o elecciones se verían afectadas. Si todo fuese verdadero, no tendríamos razones para actuar de una manera y no de otra o para elegir esto en vez de aquello. Estas consecuencias hacen que sea *absurdo* considerar la posibilidad de que todas las contradicciones (y, por tanto, todas las oraciones, etc.) sean verdaderas. Si unimos a esto el hecho de que *no tenemos razones positivas* de peso (indicios o pruebas) para aceptar la verdad de todas las contradicciones, concluiremos que la tesis de que todo es verdadero, a la que denominaremos *trivialismo*, es absurda e inaceptable.[22]

Una de las razones que habitualmente se aduce a favor del Principio de No Contradicción es que si una contradicción fuese verdadera, entonces podríamos demostrar *cualquier* cosa de donde se seguiría la verdad del trivialismo. Esta tesis recibe el nombre de Principio de *Ex Contradictione Quodlibet* (ECQ) o de *Explosión* en la terminología de Priest. De hecho, es habitual equiparar la consistencia de un sistema formal con la tesis de que "no todo es demostrable en dicho sistema", lo que presupone la validez de ECQ y la equivalencia entre "demostrar una contradicción" y "demostrar cualquier cosa". Ahora bien, ¿por qué hemos de aceptar ECQ? Una manera de justificar este principio sería afirmar que *aceptar una contradicción entraña aceptarlas todas*, tesis que desemboca

[22] Priest (2000b) señala que históricamente nadie ha prestado tanta atención a esta posición como a *su dual*, la afirmación de que todo es falso, relacionada con ciertas formas de escepticismo (aunque apunta que Aristóteles acusó a Heráclito y Protágoras de trivialismo). Priest muestra que, si tomamos los argumentos del trivialista en serio, puede ser un rival tan difícil de batir como el escéptico. En particular, Priest (ibíd., pp. 191-5) considera las objeciones al trivialismo aquí expuestas, otorgando un peso especial (aunque las razones de esta asimetría no convencen) a la imposibilidad trivialista de justificar las elecciones que hacemos. Identificar el escepticismo con la tesis dual del trivialismo (i.e., todo es falso), no resulta, sin embargo, muy plausible. Parece más adecuado describir al escéptico radical como a alguien que afirma que "nuestras creencias en la verdad de cualquier enunciado no están justificadas, por tanto, no sabemos si es verdad". Aunque también sería posible convertir al trivialista en un relativista radical según el cual "tenemos razones para creer en la verdad de todo". En cualquier caso, la idea de que el "trivialista" y el "nihilista" (i.e., el que afirma la falsedad de todo), tienen una fuerza pareja parece correcta. Después de todo, ambas posturas podrían ser equivalentes: el primero afirma la verdad (entre otras) de 'Todo es falso'; y el segundo, la falsedad (entre otras) de 'Todo es verdad'. Otro tanto cabría decir del escéptico radical (todo lo que creemos verdadero podría ser falso) y del relativista radical (todo lo que creemos falso podría ser verdadero).

en el trivialismo y en la posibilidad de demostrar la verdad de cualquier cosa.

Pero, ¿por qué aceptar *algunas* contradicciones habría de significar aceptarlas *todas*? No hay ninguna razón por la que aceptar la verdad de algunas contradicciones basándonos en los indicios racionales que aporta el examen de las paradojas nos obligue a aceptar otras que carecen de toda justificación. Según Priest, tenemos razones para aceptar que μ, la oración del mentiroso, es verdadera, pero no para aceptar $2 + 2 = 4$ y $2 + 2 \neq 4$, así pues, ¿por qué aceptar lo segundo? ¿En qué sentido implica μ la segunda contradicción? Priest rechaza como absurda la idea de que aceptar *algunas* contradicciones implique aceptarlas *todas*; y, en general, la idea de que aceptar *una* contradicción tenga necesariamente consecuencias que afecten a la *totalidad* de oraciones, enunciados o proposiciones de una teoría (i.e., que debamos aceptarlo *todo*, o que no podamos rechazar *nada*, etcétera).[23] Cualquier justificación de ECQ o PNC que presuponga tesis de este tipo debe ser, por tanto, errónea y, dado que la mayoría de argumentos a favor de PNC comete errores de este tipo según Priest (incluso los de Aristóteles en la *Metafísica*, véase Priest 1998a; y 1998b, p. 417), la mayoría de ellos se pueden rechazar directamente. Entre los argumentos que incurren en esta falta

[23] "That a person may sometimes be able to accept a contradiction rationally [...] I do not dispute. That a person can always accept a contradiction rationally, is a blatant *non sequitur*, which I reject. It does not follow from the fact that some contradictions are rationally acceptable that all are" (*IC*, p. 129). Y, de nuevo: "It is often said by way of objection to dialetheism that if one is a dialetheist, one might as well be a trivialist. This is a silly objection as are most slides from 'some' to 'all'. (Compare: you believe some things to be true; why don't you believe all things to be true?)" (Priest 2000b, p. 189). Véase también la nota 21.

encontramos los siguientes: (1) α es significativa sólo si *excluye* algo. Ahora bien, si α no excluye $\neg\alpha$, entonces *no excluye nada*, por tanto, si las contradicciones pudiesen ser verdaderas, *nada* tendría significado (*IC*, pp. 117-9; y Priest 1998b, p.418). (2) Si las contradicciones fuesen aceptables, no podríamos criticar las creencias de *nadie* racionalmente, puesto que la persona en cuestión podría aceptar sus creencias y nuestra negación de las mismas (*IC*, pp. 128-9; Priest 1998b, p. 422). (3) Si las contradicciones fuesen verdaderas, no podríamos rechazar la verdad de *nada*, ya que cuando alguien negase α afirmando $\neg\alpha$, no descartaría α por ello (*IC*, 129-32; Priest 1998b, pp. 425). Todos estos argumentos presumen erróneamente que nuestra incapacidad *en algunos casos* para rechazar α al afirmar $\neg\alpha$ implica que *nunca* podemos rechazar α al aceptar $\neg\alpha$ y, por tanto, que no podemos rechazar *nada* o, alternativamente, que debemos aceptarlo *todo*.[24]

Priest considera más argumentos a favor de PNC, entre ellos, que las contradicciones *no tienen significado*, lo que parece obviamente falso,

[24] Priest ofrece otras razones para rechazar los argumentos (1)-(3). Contra (1), afirma que es falso que para que una oración sea significativa deba *excluir* algo, la oración 'Todo es verdadero' no excluye nada y, no obstante, tiene significado (Priest 1998b, p. 418; 1999, pp. 115-6). Por otra parte, es posible entender el significado de una oración α en términos de la *información* que aporta y no de lo que α excluye. El contenido de α puede identificarse con las oraciones que α entraña [*entails*] o con $<W_1, W_2>$ (W_1 = conjunto de mundos donde α es verdadera; W_2 = mundos donde lo es $\neg\alpha$) (*IC*, pp. 118-9). Contra (2) y (3), Priest alega que las probabilidades empíricamente contrastadas de tropezar con una dialetheia son exiguas (aunque no nulas) y, por tanto, es *razonable* rechazar la inmensa mayoría de contradicciones y otros absurdos que encontremos. En caso de duda, podemos comparar las posibilidades abiertas que tenemos y decidir, de forma independiente, cuál tiene más méritos a su favor: apoyo empírico, compatibilidad con principios intuitivamente válidos y otros criterios de aceptación racional (*IC*, pp. 127, 128-32).

ya que una contradicción es la conjunción presumiblemente significativa de dos enunciados significativos. Por otra parte, ECQ presupone que las contradicciones son significativas desde el momento en que las utiliza para extraer conclusiones en un discurso significativo (ibíd., p. 417). Otro argumento a favor de PNC se basa en la imposibilidad de creer racionalmente una contradicción (*IC*, pp. 123-4). Frente a esta objeción Priest responde siempre que, en la medida en que ciertas contradicciones *son* verdaderas, es racional creer en ellas (*IC*, p. 124); así mismo, apela frecuentemente a la paradoja del prólogo para mostrar que no es tan complicado tener creencias racionales contradictorias (*IC*, p. 124; Priest 1999, p. 114).[25] Por último se puede alegar que no existe evidencia inductiva a favor de la verdad de una contradicción, pero eso no es cierto ya que las paradojas de inclusión son contradicciones verdaderas, según Priest. Lo correcto es afirmar que, inductivamente, las contradicciones verdaderas son altamente improbables, pero eso no significa que su probabilidad sea 0 (Priest 1998b, pp. 418-9; *IC*, pp. 127, 132, 134-6).

Así pues, ninguno de los argumentos a favor de PNC o de ECQ es realmente concluyente. En el caso de ECQ, Priest (1998b, pp. 411-2)

[25] Según la paradoja del prólogo, alguien podría tener buenas razones para creer conjuntamente y por separado todas las afirmaciones: $\alpha_0, \alpha_1, ..., \alpha_n$ que ha hecho en un libro tras efectuar una investigación meticulosa. Sin embargo, esa misma persona tendría razones empíricas independientes y muy poderosas para creer que todo libro, y en particular el suyo, contiene afirmaciones falsas. Esa persona creería a la vez: '$\alpha_0 \wedge \alpha_1 \wedge ... \wedge \alpha_n$' y '$\neg\alpha_0 \vee \neg\alpha_1 \vee ... \vee \neg\alpha_n$'. Priest (1998b, p. 411) utiliza también esta paradoja, para cuestionar la certeza de que nuestras creencias estén "cerradas" bajo inferencia: Podemos creer cosas de las cuales derivamos contradicciones y, sin embargo, no aceptar las contradicciones que de ellas se derivan. Así mismo (Priest 1993, p. 38), utiliza la paradoja para cuestionar que aceptar $\neg\, p$ sea equivalente a rechazar p, ya que a veces aceptamos $\neg\, p$ sin rechazar por ello p.

advierte además que, a diferencia de PNC, no ha gozado históricamente del apoyo que se le presupone hoy en día. En particular, Aristóteles no considera su validez, no existen pruebas reales de que fuese aceptado por los estoicos y no aparece claramente formulado hasta la Edad Media (Guillermo de Soissons), donde fue defendido por los Parvipontinianos y posteriormente aceptado por Escoto. También entonces fue un principio polémico, la escuela de Colonia lo cuestionó en el siglo XV. ECQ sólo se encumbró a finales del XIX y principios del XX con las obras de Boole, Frege y los padres de la lógica moderna y debe su prestigio, en parte, a este origen. Sin embargo, Priest no es el único que ha cuestionado ECQ, todas las lógicas paraconsistentes nacidas en la segunda mitad del siglo XX lo rechazan.

Pese a todo, no es cierto que carezcamos de argumentos independientes para aceptar la validez de ECQ. Aunque, eso sí, para defender ECQ necesitamos defender una tesis a la que se apela a menudo para *justificar* PNC, a saber, la tesis de que "*toda contradicción es absurda*". Esta tesis puede representarse del siguiente modo: '$(\alpha \wedge \neg\alpha) \rightarrow \perp$' (donde '$\perp$' es un símbolo equivalente a un absurdo lógico).[26] A partir de esta tesis y de una regla de inferencia fundamental: *Reductio ad Absurdum* (RA), se puede demostrar fácilmente una fórmula: '$(\alpha \wedge \neg\alpha) \rightarrow \beta$' cuyo contenido intuitivo se corresponde con ECQ. Por "*Reductio*

[26] De hecho, solemos aceptar una afirmación más fuerte: $(\alpha \wedge \neg\alpha) \leftrightarrow \perp$. Pero, pese a ser válida en la lógica clásica y en la intuicionista, esta afirmación es demasiado fuerte, ya que no es válida en todas las lógicas consistentes. La lógica minimal es consistente, pero rechaza: $\perp \rightarrow (\alpha \wedge \neg\alpha)$. Para nuestros propósitos (i.e., justificar ECQ) sólo necesitamos aceptar $(\alpha \wedge \neg\alpha) \rightarrow \perp$ (que sí es válida en todas las lógicas consistentes, incluida la minimal).

ad Absurdum" entendemos aquí el siguiente principio: "Si la suposición de α nos lleva a un *absurdo*, entonces podemos negar/rechazar α". RA puede representarse mediante la fórmula '$(\alpha \rightarrow \bot) \rightarrow \neg\alpha$'. Si aceptamos que *toda* contradicción es absurda, entonces tenemos '$(\alpha \wedge \neg\alpha) \rightarrow \bot$' e inferimos '$\neg(\alpha \wedge \neg\alpha)$' (i.e., PNC) apelando a RA. También podemos inferir ECQ apelando a RA y a otras reglas clásicas de inferencia muy básicas.[27] RA es un principio de gran importancia que utilizamos, por ejemplo, para explicar cómo se introduce la negación. Bajo esta interpretación, RA se convierte en una regla básica: $[\alpha], ..., \bot \Rightarrow \neg\alpha$, donde $[\alpha]$ indica que hemos supuesto α y los puntos suspensivos abrevian los pasos que nos llevan a $\bot$, un absurdo a partir del cual inferimos la negación de α.[28] Independientemente de cómo interpretemos

---

[27] La demostración de ECQ: $(\alpha \wedge \neg\alpha) \rightarrow \beta$, suponiendo '$(\alpha \wedge \neg\alpha) \rightarrow \bot$', es sencilla:

0. $(\alpha \wedge \neg\alpha) \rightarrow \bot$ Premisa: "*Toda contradicción es absurda.*"
1. $\alpha \wedge \neg\alpha$ (Suposición)
2. $\neg\beta$ (Suposición)
3. α ($\alpha \wedge \neg\alpha \Rightarrow \alpha$ (Eliminación '$\wedge$'): 1)
4. $\neg\alpha$ ($\alpha \wedge \neg\alpha \Rightarrow \neg\alpha$ (Eliminación '$\wedge$'): 1)
5. $\alpha \wedge \neg\alpha$ ($\alpha, \neg\alpha \Rightarrow \alpha \wedge \neg\alpha$ (Introducción '$\wedge$'): 3, 4)
6. $\bot$ (Modus Ponens: 0, 5)
7. $\neg\neg\beta$ ($[\alpha], ..., \bot \Rightarrow \neg\alpha$ (RA/Introducción '$\neg$'): 2-6)
8. β ($\neg\neg\alpha \Rightarrow \alpha$ (Doble negación/Eliminación '$\neg$'): 7)
9. $\alpha \wedge \neg\alpha \rightarrow \beta$ ($[\alpha] ... \beta \Rightarrow \alpha \rightarrow \beta$ (Introducción '$\rightarrow$'): 1-8)

Al emplear en el paso 8 la regla de doble negación "$\neg\neg\alpha \Rightarrow \alpha$", esta demostración no es válida para un intuicionista, sin embargo, para ellos ECQ es una regla *básica* (no derivada).

[28] A menudo se distingue entre las reglas de "introducción de la negación" (I$\neg$) y RA. La primera se identifica con lo que nosotros hemos llamado RA: "$[\alpha], ..., \bot \Rightarrow \neg\alpha$", mientras que se reserva el nombre de RA a: "$[\neg\alpha], ..., \bot \Rightarrow \alpha$". Si aceptamos la equivalencia: '$\alpha \leftrightarrow \neg\neg\alpha$' (como hacen el lógico clásico y el dialetheista –véase la nota 38–), ambas reglas son equivalentes. Los intuicionistas no aceptan '$\neg\neg\alpha \rightarrow \alpha$' y, por tanto, rechazan "$[\neg\alpha], ..., \bot \Rightarrow \alpha$". Para ellos la única *reductio* válida es "$[\alpha], ..., \bot \Rightarrow \neg\alpha$" y, ciertamente, ésta parece la forma más natural de expresar la idea de *reductio*. Por

RA, parece claro que, una vez rechazado PNC, tampoco RA puede ser válido (*si entendemos que toda contradicción es absurda*). Si afirmásemos que una contradicción y un absurdo son lógicamente equivalentes (como hacen la mayoría de lógicas consistentes: "$(\alpha \wedge \neg\alpha) \leftrightarrow \bot$"), lo que RA afirmaría es que, siempre que lleguemos a una contradicción, hemos de rechazar la verdad de alguna de las cosas cuya suposición nos llevó a inferirla. Pero ésta es, precisamente, la idea a la que se opone Priest y el principio al que apelan las soluciones clásicas o consistentes a las paradojas cuando cuestionan la validez general de PEV, PEA, etc. al tropezar con contradicciones.

Esto pone de manifiesto una cuestión crucial: Rechazar PNC y ECQ comporta rechazar *la validez general* de otros muchos principios de inferencia que gozan de una amplia aceptación. Uno de ellos, como ya hemos visto, es RA, pero también los principios de *Modus Ponens* (MP), *Silogismo Disyuntivo* (SD) o *Transitividad de* $\rightarrow$ (T$\rightarrow$) carecen de validez general si existen dialetheias y ECQ es inválido.[29] Estos principios presuponen la validez de PNC. Así pues, el dialetheismo se enfrenta a un problema similar (volveremos a ello en 5.3.2) al que se enfrentan las teorías consistentes: Rechazar la validez general de PNC y ECQ tiene un precio muy elevado con respecto a nuestras prácticas demostrativas, nos fuerza a revisar por completo nuestra lógica. Quine insistiría aquí en que "cambiar de lógica es cambiar de tema", pero, según

---

eso reservamos el nombre de RA para este principio y no para su principio hermano (cuya validez no cuestionamos).

[29] El dialetheismo excluye aún otras leyes clásicas, por ejemplo, $\neg\alpha \models \alpha \rightarrow \beta$ (Priest 1999, p. 110).

Priest (1998b, p. 416), esto presupone la validez de PNC, ECQ y de la lógica clásica en general, lo que incurre en una petición de principio contra el dialetheismo. El reto de Priest consiste, pues, en construir una lógica paraconsistente satisfactoria (cuestión que abordaremos en 5.2.3) y en convencernos de su conveniencia. Quizá por ello, Priest concluye que no ha demostrado la falsedad de ECQ, sino, más bien, la plausibilidad de su rechazo. Al final, todo depende de qué teorías (las consistentes o las inconsistentes) son más coherentes cuando apelamos a los factores de aceptación racional que, en su conjunto, determinan la superioridad de una teoría frente a otras.[30]

5.2.2.2 Verdadero, no verdadero, falso

Priest defiende que el dialetheismo, la tesis de que hay contradicciones verdaderas, es independiente de la noción o teoría de *verdad* que uno baraje (correspondencia, coherentismo, pragmatismo, deflacionismo, etc.).[31] Pero, empleemos la teoría que empleemos, la

[30] "I have not shown that explosion [ECQ] fails, that one ought to take into the scope of logic situations that are inconsistent and/or incomplete, though I do take it that, when the dust settles, this will be seen to be the case [...] The point [...] is simply to show that the failure of explosion is a plausible logico-metaphysical view and that one cannot simply assume otherwise without begging the question" (Priest 1998b, p. 416).

[31] "Dialetheism, the view that some contradictions are true, does not commit one *per se* to any particular account of truth." (*IC*, p. 67). Priest 2000a y Armour-Garb y Beall 2002 desarrollan argumentos a favor de esta tesis. Un dialetheista que apoyase una teoría de la verdad como correspondencia aceptaría la *existencia* de contradicciones *en* el mundo y de *hechos* inconsistentes. Sin comprometerse con tal teoría, Priest afirma que las paradojas de inclusión son contradicciones verdaderas *en* el mundo: "I claim that reality is, in a certain sense, contradictory. I do not, of course, mean that the objects that constitute reality, like chairs and stars, are contradictory. That would simply be a category mistake. What I mean is that there are certain contradictory statements (propositions, sentences –take your pick) about limits, that are true. I am enough of a

extensión de 'verdadero' debe determinarse por completo mediante PEV, el esquema tarskiano (*IC*, p. 69-74). Priest apoya, de hecho, una teoría *teleológica* de la verdad inspirada en Dummett (*IC*, pp. 77-9) según la cual la extensión de 'verdadero' está fijada por PEV, pero sólo podemos captar el *significado* de 'verdadero' ('V' para abreviar) si se nos informa además de que la verdad es el *objetivo* que persigue la gente habitualmente *al hacer aserciones*.[32] Dada la supuesta neutralidad del dialetheismo en este respecto, sólo la satisfacción de PEV nos preocupará aquí y esto por dos motivos: En primer lugar, porque PEV nos ayudará a definir el predicado 'falso' ('F' para abreviar) a partir de 'verdadero' y, en segundo, porque el bicondicional: V($\underline{p}$) $\leftrightarrow p$ ($\underline{p}$ nombra la oración –o proposición, etc.– p), debe cumplir *siempre*: (1) $p, p \rightarrow$ V($\underline{p}$) $\vdash$ V($\underline{p}$); y (2) V($\underline{p}$), V($\underline{p}$) $\rightarrow p \vdash p$. Acabamos de ver en 5.2.2.1 que esto *no* es posible si entendemos '$\rightarrow$' como un condicional *material*, ya que este condicional no satisface en la lógica de Priest la ley de *Modus Ponens*. Abordaremos esta cuestión al final de 5.2.3.1, ahora nos ocuparemos de PEV y de las nociones de "falsedad" y "no-verdad" [*untruth*].

---

realist to hold that there must be something about reality that makes them so [...]. When I say that reality is contradictory, I mean that it is such as to render those contradictory statements true." (*BLT*, p. 295).

[32] "Dummett compares asserting with playing a game, and speaking the truth with winning. [...] Playing a game has a *telos*: winning (or more precisely, obtaining a winning position). [...] We could specify extensionally what it is to have a winning position in bridge, chess and so on, but someone who knew only this would not know what winning is. What they would need to know is that winning is what people play the game to achieve. Similarly, the T-scheme [PEV] may characterise what it is for each particular sentence to be true. But unless a person knows that truth is what people who assert aim to speak, she will not know what truth is." (*IC*, p. 78). Para defender su teoría de la verdad, Priest se apoya en una teoría de la aserción inspirada en Grice que discute y defiende en *IC*, pp. 79-80.

La noción de "verdad" nos ayuda a caracterizar negativamente las nociones de "falsedad" y "no-verdad". Estas tres nociones cumplen, desde una perspectiva clásica, las siguientes condiciones: (1) Una oración (proposición, etc.), *p*, *es falsa* si, y sólo si, *su negación es verdadera*: $F(\underline{p}) \leftrightarrow V(\underline{\neg p})$ (aplicando PEV: $V(\underline{\neg p}) \leftrightarrow \neg p$, obtenemos, además, $F(\underline{p}) \leftrightarrow \neg p$). (2) Las nociones de verdad y falsedad son *exhaustivas* (i.e., *toda* oración es verdadera *o* falsa) y *mutuamente excluyentes* (i.e., *ninguna* oración es verdadera *y* falsa). (3) Las nociones de "falsedad" y "no-verdad" son *coextensivas*: $F(\underline{p}) \leftrightarrow \neg V(\underline{p})$. (4) Por consiguiente, "verdad" y "no-verdad" *también* son nociones exhaustivas y excluyentes.

Comencemos por la noción de falsedad. Al igual que los lógicos clásicos, los dialetheistas aceptan las equivalencias de (1) en su definición de 'falso' (*IC*, pp. 81, 100). Sin embargo, cualquier dialetheista debe rechazar (3), ya que las nociones de verdad y falsedad *no son* mutuamente excluyentes, la verdad de una oración (proposición, etc.) no excluye su falsedad, y viceversa (*IC*, p. 84): Una paradoja de inclusión, *p*, cumple $F(\underline{p}) \wedge V(\underline{p})$. Un dialetheista puede aceptar, en cambio, que las nociones de verdad y falsedad sean exhaustivas: $V(x) \vee F(x)$ (para toda oración, proposición, etc., x) y, por tanto, que si x no es verdadera, deba ser falsa: $\neg V(x) \rightarrow F(x)$, y viceversa: $\neg F(x) \rightarrow V(x)$.[33] Priest señala que aceptar esta postura es independiente del dialetheismo: Un dialetheista podría negar que 'V' y 'F' fuesen exhaustivos (*IC*, p. 83)

[33] Como la ley de Silogismo Disyuntivo (SD) no es válida en general y, por tanto, el condicional material no satisface la ley de *Modus ponens* (MP), el signo '$\rightarrow$' en: '$\neg V(\underline{p}) \rightarrow F(\underline{p})$' y '$\neg F(\underline{p}) \rightarrow V(\underline{p})$' se debe considerar un operador de implicación no material que cumple MP, véase el final de 5.2.3.1.

si compartiese los argumentos que esgrime el intuicionista contra PTE o los argumentos de quienes apoyan la existencia de oraciones sin valor de verdad debido a fallos semánticos: oraciones no fundamentadas, presuposiciones no satisfechas, presencia de nombres "vacíos", etcétera. Si ellos tuviesen razón, $V(\underline{p})$ y $F(\underline{p})$ podrían *ambas* carecer de justificación adecuada en ciertas circunstancias. No obstante, Priest rechaza esta posibilidad basándose (no en sus convicciones dialetheistas sino) en su teoría teleológica de la verdad (*IC*, p. 81-2).[34] Según Priest, la aserción es un juego individual en el que no puede haber "tablas": el objetivo es la verdad, si el objetivo no se consigue, lo aseverado es falso. Su argumento principal consiste en afirmar que, si no hay *nada* en el mundo que justifique la verdad (falsedad) de *p*, entonces *esa ausencia* de pruebas o hechos a favor de $V(\underline{p})$ ($F(\underline{p})$) constituye en sí un *hecho* –al que llama "negativo"– a favor de la falsedad (verdad) de *p*, justificando así $F(\underline{p})$ ($V(\underline{p})$).[35] Desde esta perspectiva, la oración del veraz, $\underline{v}$ = '$V(\underline{v})$' (que afirma su propia verdad), y otras similares serían falsas, ya que *es un hecho* ("negativo") que *ningún hecho* del mundo avala su verdad, lo que la convierte en falsa (*IC*, p. 84).

[34] Tanto su teoría teleológica de la verdad (*IC*, pp. 76-9) como el argumento en contra de la exhaustividad de 'V' y 'F' (*IC*, pp. 81-2) se inspiran en Dummett, lo que no deja de ser curioso porque la mayoría de argumentos intuicionistas considerados contra el principio de tercio excluso y la "exhaustividad" de 'V' y 'F' se inspiran *también* en Dummett (véase, por ejemplo, Priest 1999, p. 105).

[35] "If there is no Fact that makes α true, there is a Fact that makes $\neg\alpha$ true, viz. the Fact that there is no Fact that makes α true. Less cryptically the point is this. Suppose that α is a sentence, and suppose that there is nothing in the world in virtue of which α is true; no fact, no proof, no experimental test. Then this is the Fact in virtue of which $\neg\alpha$ is true. We may not know this Fact obtains, but this is irrelevant." (*IC*, p. 83)

Esta maniobra fuerza una distinción entre hechos "negativos" y "positivos" (*IC*, p. 85). Una oración, *p*, falsa en virtud de hechos *negativos* (i.e., *ausencia* de hechos positivos a favor de su verdad) no puede ser además verdadera, ya que eso implicaría la *existencia de hechos positivos* a favor de su verdad. Así pues, una dialetheia, i.e., una oración (proposición, etc.), *p*, *verdadera y falsa*, debe ser ambas cosas *en virtud de hechos positivos* a favor de $V(\underline{p})$ y de $F(\underline{p})$. En el caso de las paradojas de inclusión, las pruebas de trascendencia y clausura aportan estos hechos positivos. Priest considera, además, otros casos que involucran factores empíricos en los que dos tests diferentes, pero igualmente válidos (y habitualmente equivalentes) a la hora de clasificar objetos de cierto tipo (partidos políticos de izquierda o derecha; temperaturas por encima o por debajo de *x* grados), ofrecen resultados contradictorios (*IC*, pp. 85-7).

Hasta ahora hemos hablado de "falsedad", pero ¿es *ser falso* lo mismo que *no ser verdadero*? Para que ambas nociones coincidan deben cumplirse dos cosas: $\neg V(\underline{p}) \rightarrow F(\underline{p})$ y $F(\underline{p}) \rightarrow \neg V(\underline{p})$. Priest acepta el primer condicional por ser 'V' y 'F' exhaustivos (recordemos, sin embargo, que esta tesis es independiente del dialetheismo). El segundo condicional supone, en cambio, que 'V' y 'F' son excluyentes y es, por tanto, rechazado.[36]

[36] Priest considera la posibilidad de que $F(\underline{p}) \rightarrow \neg V(\underline{p})$ se cumpla. Esto entraría en contradicción con la tesis dialetheista de que 'V' y 'F' *no* son excluyentes. Para Priest ésta no es razón suficiente para rechazar $F(\underline{p}) \rightarrow \neg V(\underline{p})$, ya que lo que persigue un dialetheista es, precisamente, acomodar ciertas contradicciones (*IC*, p. 88). Pese a todo, Priest rechaza $F(\underline{p}) \rightarrow \neg V(\underline{p})$ por considerar: (1) que las razones que apoyan su verdad

Para el dialetheista las diferencias entre falsedad y no-verdad son mínimas (también verdad y no-verdad son exhaustivas y no excluyentes), pero existen. Falsedad y no-verdad *no son* nociones coextensivas. Priest ilustra este hecho mediante la paradoja reforzada del mentiroso: $\underline{\mu}$ = '$\neg$V($\underline{\mu}$)'. Según Priest (*IC*, pp. 90-1), podemos demostrar a partir de PEV, μ y la ley de tercio excluso (que él ha aceptado por ser 'V' y 'no-V' exhaustivos): $\exists x(\mathrm{V}(x) \wedge \neg\mathrm{V}(x))$ y $\neg\exists x(\mathrm{V}(x) \wedge \neg\mathrm{V}(x))$. Sin embargo, ninguna oración (incluida μ o la del mentiroso) justifica: $\exists x(\mathrm{V}(x) \wedge \mathrm{F}(x))$ y $\neg\exists x(\mathrm{V}(x) \wedge \mathrm{F}(x))$. Las nociones de "verdad" y "no-verdad" son *a la vez* mutuamente excluyentes y no excluyentes y, por tanto, más inconsistentes que las nociones de "verdad" y "falsedad" que, sencillamente, *no son* mutuamente excluyentes.

5.2.2.3 La negación

Al definir la noción de "falsedad", hemos dicho en 5.2.2.2 que una oración, *p*, es falsa si, y sólo si, su negación es verdadera: $\mathrm{F}(\underline{p}) \leftrightarrow \mathrm{V}(\underline{\neg\, p})$. Ahora bien, ¿cómo definir la negación? Priest apunta que las nociones de "falsedad" y "negación" son interdependientes, podemos definir "falsedad" en términos de "verdad" y "negación", pero también podemos definir "negación" en términos de "verdad" y "falsedad" (i.e., *el operador que convierte una oración verdadera en falsa y viceversa*). Así pues, todo intento de ofrecer una definición conjunta es circular (*IC*, p. 81), o bien tomamos una como básica e indefinible, o bien tomamos

no son concluyentes (rechaza la contraposición de PEV: $\neg\mathrm{V}(\underline{p}) \leftrightarrow \neg p$); y (2) que es conveniente *no multiplicar contradicciones innecesariamente* (*IC*, pp. 89-90).

ambas como indefinibles y exploramos cómo se relacionan entre sí. Pese a los problemas de circularidad mencionados, son muchas las cosas relevantes que se pueden decir sobre la negación (véase Priest 1999):[37] Una negación es una operación que, por así decirlo, cambia la "polaridad" de aquello sobre lo que se aplica generando pares de elementos *contradictorios* que representamos mediante α y $\neg\alpha$. Hablar de negación es hablar de la *relación de contradicción* que mantienen entre sí α y $\neg\alpha$ (Priest, 1999, p. 103). Esta relación no debe confundirse con las relaciones de *contrariedad* y de *sub-contrariedad.* Una oración α podría tener mas de un contrario o sub-contrario pero sólo un elemento contradictorio: $\neg\alpha$ es único. Además, la relación de contradicción satisface (a diferencia de las otras dos) el siguiente requisito definitorio: De los dos elementos contradictorios, α y $\neg\alpha$, *al menos* y *a lo sumo uno* debe ser el caso. De este requisito se sigue la validez de dos principios: el de tercio excluso (PTE): $\alpha \vee \neg\alpha$; y el de no contradicción (PNC): $\neg(\alpha \wedge \neg\alpha)$ (ibíd., p. 104). La relación de contradicción es, además, simétrica, por lo que el contradictorio del contradictorio de α, i.e., $\neg\neg\alpha$, es el propio α, de donde se sigue el principio de doble negación (PDN): $\neg\neg\alpha \leftrightarrow \alpha$ (ibíd., pp. 106-7).[38]

[37] Priest (1999, p. 103) insiste en que estudiar la *negación* no equivale a estudiar el uso que hacemos en los lenguajes naturales de la palabra 'no' ('*not*', etc.), ya que muchas veces *negamos* usando otras palabras ('ningún', etc.) y encontramos oraciones que, pese a contener un 'no' y tener forma negativa, no pretenden negar aquello sobre lo que se aplican sino mostrar otro tipo de rechazo ("*No* soy su mujer, él es mi marido").

[38] "We see that there appears to be a relationship of some kind between pairs such as 'Socrates is mortal' and 'Socrates is not mortal'; and 'Some man is mortal' and 'No man is mortal'. The traditional way of expressing the relationship is that the pairs are *contradictories*, and so we may say that the relationship is that of contradiction.

Priest acepta PTE apoyándose en el mismo tipo de razones que le llevan a aceptar la tesis de que toda oración (proposición, etc.) es verdadera o falsa. Priest aduce además que la noción intuicionista de negación (que rechaza PTE) no es apta para formar pares de contradictorios y, por tanto, no puede ofrecer una teoría correcta de la negación (ibíd., pp. 104-5). Lo que resulta más sorprendente, en cambio, es que, en cierto sentido, Priest *también acepta* PNC. En su lógica dialetheista $\vdash \neg(\alpha \wedge \neg\alpha)$ es un teorema válido para cualquier oración (proposición, etc.) α. Ahora bien, si α es una paradoja de inclusión, entonces tenemos: $\vdash \alpha \wedge \neg\alpha$ (por las pruebas de trascendencia y

---

Theories of negation are theories about this relation." (Priest, 1999, p. 103). Priest caracteriza la relación de contradicción: "So if α is any statement, let $\neg\alpha$ represent its contradictory. (Contradictories, unlike contraries and sub-contraries are unique –at least up to logical equivalence.) What relationship holds between these? Traditional logic and common sense are both very clear about the most important one: we must have at least one of the pair but not both. It is precisely this which distinguishes contradictories from their near cousins, contraries and sub-contraries. [...] This fact about contradictories gives immediately two of the traditional laws of negation, the law of excluded middle (LEM), $\alpha \vee \neg\alpha$, and the law on non-contradiction (LNC), $\neg(\alpha \wedge \neg\alpha)$. [...] Now, maybe the traditional claim about contradictories –and consequently these two laws– is wrong; but it would certainly be the default position. The onus of proof is therefore on those who would dispute it." (Ibíd., p. 104). De estas leyes derivamos, según Priest, la de doble negación: "[T]he law of double negation (LDN) is simply derivable. The relationship of being contradictories is symmetric. If β is the contradictory of α, then α is the contradictory of β. In particular, α is the contradictory of $\neg\alpha$. Hence, $\neg\neg\alpha$ just is α." (Ibíd., pp. 106-7). No está claro, sin embargo, que podamos derivar LDN a partir de PTE y PNC prescindiendo de leyes, como la de *Silogismo Disyuntivo* (SD) o *Reductio ad Absurdum* (RA), que son inválidas para el dialetheista. La derivación más natural de LDN ($\neg\neg\alpha \leftrightarrow \alpha$) a partir de PNC y PTE parece la siguiente: Supongamos $\neg\neg\alpha$, sabemos que $\alpha \vee \neg\alpha$ (por PTE), de donde inferimos α (*por SD*). Así pues, $\neg\neg\alpha \rightarrow \alpha$. Supongamos ahora α, si tuviésemos $\neg\alpha$, inferiríamos $\alpha \wedge \neg\alpha$ (por PNC), así pues, $\neg\neg\alpha$ (*por RA*), por tanto, $\alpha \rightarrow \neg\neg\alpha$, de donde concluimos: $\neg\neg\alpha \leftrightarrow \alpha$.

clausura); y $\vdash \neg(\alpha \wedge \neg\alpha)$ (por PNC).[39] Esta contradicción no inquieta a Priest por dos motivos, en primer lugar, porque el dialetheismo defiende, precisamente, la verdad de algunas contradicciones y, en segundo, porque esta contradicción "de segundo orden" *no es arbitraria*, se deriva a su vez de una contradicción del tipo que genera paradojas de inclusión autorreferentes. Alguien podría objetar que esta contradicción de orden superior es de una naturaleza diferente a las otras (a la del mentiroso por ejemplo) ya que se produce en el *metalenguaje* en el que Priest construye su teoría dialetheista, dando a entender que aceptamos y rechazamos a la vez PNC. Sin embargo, para Priest, esta objeción no es válida ya que una de las motivaciones de su teoría es evitar la distinción entre lenguaje y metalenguaje que trazaban las soluciones clásicas a la paradoja del mentiroso. Para Priest esta distinción está injustificada y es problemática, sólo existe un lenguaje y es, efectivamente, inconsistente.[40] Priest insiste, además, en que este tipo de contradicciones "de segundo orden" derivan siempre de enunciados paradójicos reconocidos como tales (*IC*, p. 91).

Priest (1999, p. 108) distingue entre PTE: $\alpha \vee \neg\alpha$; y el principio de "bivalencia": $\forall x(V(x) \vee F(x))$, donde x es una variable a sustituir por oraciones (o proposiciones, etc.). También distingue entre PNC: $\neg(\alpha \wedge \neg\alpha)$; y el principio de "consistencia": $\forall x\neg(V(x) \wedge F(x))$. Priest acepta tanto PTE como el principio de bivalencia y, como acabamos de ver, acepta PNC y la negación de PNC, sin embargo rechaza de plano el

[39] "[T]he fact that that every instance of $\neg(\alpha \wedge \neg\alpha)$ is true does not, on its own, prevent some instances of α and $\neg\alpha$ from being true." (Priest, 1999, p. 108).
[40] Esta idea es recurrente, véase Priest 1979, pp. 238-40; 1999, pp. 111-2. Véase notas 36, 80 y 92.

principio de consistencia. Esto está relacionado con el hecho apuntado al final de 5.2.2.2: existen motivos para aceptar $\exists x(V(x) \wedge F(x))$, pero no para aceptar $\neg\exists x(V(x) \wedge F(x))$. Así pues, el rechazo del principio de consistencia no es en sí mismo contradictorio para un dialetheista, a diferencia de su rechazo de PNC.

Es frecuente distinguir entre las nociones de negación y "rechazo" [*denial*]. Negar una oración es una operación de carácter eminentemente semántico, rechazarla tiene connotaciones pragmáticas. La negación de α es su contradictorio $\neg\alpha$, una oración cuyo *contenido semántico* difiere de α. Negar α es, pues, aceptar $\neg\alpha$. Sin embargo, podemos rechazar α pese a aceptar su contenido semántico (i.e. sin comprometernos con $\neg\alpha$) por ser α inconveniente en algún sentido pragmático: irrelevante, engañoso, etc. (véase nota 37). Una forma habitual (aunque no la única) de expresar nuestro rechazo de α es afirmar su negación: $\neg\alpha$. La lógica clásica defiende que negar una oración (proposición, etc.) *implica siempre* rechazarla: afirmar $\neg\alpha$ es rechazar α. Priest cuestiona esta tesis, un dialetheista puede afirmar $\neg\alpha$ sin rechazar por ello α cuando se halla en presencia de una paradoja de inclusión. Pero, dejando a un lado el dialetheismo, Priest (1999, p. 114) ofrece otros argumentos. Muchas veces aceptamos contradicciones de forma inconsciente (sólo cuando reparamos en ellas revisamos nuestras creencias), en estos casos, afirmar $\neg\alpha$ no puede constituir inmediatamente un rechazo de α. Pero hay incluso casos, la paradoja del prólogo expuesta en la nota 25 (5.2.2.1), en los que puede ser razonable aceptar creencias contradictorias. Priest admite que la negación expresa a menudo rechazo, sólo se opone a la tesis general de que *siempre* exprese rechazo. En algunos contextos (que no podemos determinar de antemano) negar algo implica rechazarlo, en otros no.

5.2.3 Lógica paraconsistente (LP) y modelos inconsistentes

Describiremos en esta sección la semántica y los elementos básicos de LP, la lógica paraconsistente que diseña Priest, junto con algunos resultados relativos a modelos inconsistentes. Señalaremos cuáles son los principales rasgos de LP, ubicando en la bibliografía los textos en los que Priest obtiene los resultados formales más relevantes. Para obtener más información se remitirá al lector a las obras aquí citadas.

5.2.3.1 La semántica de LP[41]

Priest parte de un lenguaje de primer orden, LP, con las conectivas lógicas tradicionales ($\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\forall$ y $\exists$) que contiene variables de individuo, constantes y predicados (incluido el de identidad, 'verdadero' y 'falso' –los dos últimos entendidos como predicados parcialmente independientes–), y que también puede contener símbolos de función.[42] Un modelo M de LP es un par ordenado, $M = \langle I, D\rangle$, donde D es el dominio de interpretación e I la función que interpreta LP a partir

[41] Véase Priest 1979, III, pp. 226-31; *IC*, capítulo 5, pp. 94-8; o Priest 1991, pp. 323-4. El resumen más conciso de LP y sus propiedades se puede encontrar en *BLT*, pp. 171-2.

[42] Priest (1991, p. 323; *BLT*, p. 171) apunta que la inclusión de símbolos de función complica (aunque no imposibilita) la demostración de ciertos resultados formales de particular importancia para LP, sobre todo el *Collapsing Lemma* (véase 5.2.3.3). Siguiendo al propio Priest, excluiremos de LP estos símbolos para simplificar las cosas.

de D. I asigna un elemento de D a cada constante 'c' de LP: $I(\mathrm{c}) \in D$. LP tiene, al menos, un nombre (una constante), 'c', para cada elemento, c, de D. A cada predicado n-ario, 'P', de LP, I le asigna un par ordenado: $I(\mathrm{P}) = \langle I^{+}(\mathrm{P}), I^{-}(\mathrm{P})\rangle$, que consta de $I^{+}(\mathrm{P}) \subseteq D^{\mathrm{n}}$ (la extensión de 'P' en D), e $I^{-}(\mathrm{P}) \subseteq D^{\mathrm{n}}$ (su antiextensión). Los predicados de LP cumplen la ley de tercio excluso ($I^{+}(\mathrm{P})$ e $I^{-}(\mathrm{P})$ son "exhaustivos"): $I^{+}(\mathrm{P}) \cup I^{-}(\mathrm{P}) = D^{\mathrm{n}}$; sin embargo, no son necesariamente consistentes, existen predicados (como 'V', i.e., *verdadero*) tales que: $I^{+}(\mathrm{P}) \cap I^{-}(\mathrm{P}) \neq \varnothing$. La interpretación de los términos de LP se completa con una función, s, que asigna elementos de D a sus variables: $s(x_{\mathrm{n}}) \in D$. Definiremos $I^{*} = I \cup s$ como la función que asigna a cada término, $t \in$ LP, su *denotación* en D: Si t es una constante, $I^{*}(t) = I(t)$; si t es una variable, $I^{*}(t) = s(t)$. Al aceptar la posibilidad de que una oración $p \in$ LP sea verdadera y falsa a la vez, la interpretación semántica de las *fórmulas* y *conectivas* de L no puede ser clásica. Desde un punto de vista extensional, toda oración p puede recibir *dos valores de verdad*: verdadero (1) o falso (0), sin embargo, caben *tres posibilidades evaluativas*: o p es 1; o p es 0; o p es 1 y 0. Hemos visto que los *valores de verdad* no son excluyentes, una oración puede ser falsa *y* verdadera a la vez, sin embargo, las *posibilidades evaluativas* sí lo son: una oración debería ser *o sólo* verdadera, *o sólo* falsa, *o* verdadera *y* falsa.

Combinar dos valores de verdad con tres posibilidades evaluativas introduce una complicación añadida. La evaluación de cada fórmula no puede llevarse a cabo mediante una *función* sino mediante una *relación*. Una dialetheia debería recibir *dos* valores de verdad, pero, por definición, una función *sólo* puede asignar *un único* valor de verdad a cada fórmula.

Una evaluación es, por tanto, una relación, *R*, que asigna *uno o dos* valores de verdad a cada fórmula de LP, incluidas las oraciones (i.e., fórmulas sin variables libres). En el caso de una fórmula atómica (con términos $t_0, ..., t_n$): $\mathrm{P}(t_0, ..., t_n)$, *R* operaría del siguiente modo:

$R(\mathrm{P}(t_0, ..., t_n), 1)$ si, y sólo si, $(I^*(t_0), ..., I^*(t_n)) \in I^+(\mathrm{P}) \subseteq D^n$

$R(\mathrm{P}(t_0, ..., t_n), 0)$ si, y sólo si, $(I^*(t_0), ..., I^*(t_n)) \in I^-(\mathrm{P}) \subseteq D^n$ [43]

Si una oración, *p*, es una dialetheia, entonces tenemos $R(p, 1)$ y $R(p, 0)$. Aunque Priest parece primar esta concepción *relacional* de las evaluaciones, ofrece también una presentación *funcional* alternativa para simplificar y hacer más manejable la semántica de LP.[44] Priest define las evaluaciones de oraciones mediante una *función*, *v*, que asigna a cada fórmula de LP un elemento del conjunto $\{\{0\}, \{1\}, \{0, 1\}\}$. Desde esta perspectiva, los valores de verdad son "técnicamente" tres, aunque los únicos valores genuinos que acepta la teoría sean sólo dos: verdadero y falso. Una oración *p* es verdadera si, y sólo si ('sii' en lo sucesivo), 1 ∈

---

[43] Las conectivas '$\neg$', '$\wedge$' y '$\forall$', por ejemplo, se define del siguiente modo:

$R(\neg p, 1)$ si, y sólo si, $R(p, 0)$ $\quad$ $R(p \wedge q, 1)$ si, y sólo si, $R(p, 1)$ y $R(q, 1)$

$R(\neg p, 0)$ si, y sólo si, $R(p, 1)$ $\quad$ $R(p \wedge q, 0)$ si, y sólo si, $R(p, 0)$ o $R(q, 0)$

Dada una fórmula F con una variable libre, *x* (*x*/*a* es el resultado de sustituir *x* por la constante 'a'):

$R(\forall x\mathrm{F}, 1)$ si, y sólo si, $R(\mathrm{F}(x/\mathrm{a}), 1)$ para toda constante a ∈ LP (i.e., para todo $a \in D$, ya que $I(\mathrm{a}) = a$)

$R(\forall x\mathrm{F}, 0)$ si, y sólo si, $R(\mathrm{F}(x/\mathrm{a}), 0)$ para alguna constante a ∈ LP (i.e., para un $a \in D$, ya que $I(\mathrm{a}) = a$)

[44] "[W]e can conceive of an evaluation as a subset of $\mathrm{F} \times \{0, 1\}$ [F es el conjunto de fórmulas de LP], where every member of F occurs as the first member of at least one pair in the evaluation. Actually, it is technically simpler to think of an evaluation, equivalently, as a function which maps a formula to the set of truth values [a saber, el conjunto: $\{\{0\}, \{1\}, \{0, 1\}\}$] to which the formula is related, so this is how we will proceed." *IC*, p. 94. La concepción *relacional* de las evaluaciones es, sin embargo, la preferida, véase Priest 1993, pp. 50-1; y Priest 1998b, p. 413.

$v(p)$; y falsa sii $0 \in v(p)$; p es una dialetheia sii $0, 1 \in v(p)$. Priest describe recursivamente la semántica de LP:

$1 \in v(\mathrm{P}(t_0, ..., t_n))$ sii $(I^*(t_0), ..., I^*(t_n)) \in I^+(\mathrm{P}) \subseteq D^n$

$0 \in v(\mathrm{P}(t_0, ..., t_n))$ sii $(I^*(t_0), ..., I^*(t_n)) \in I^-(\mathrm{P}) \subseteq D^n$

$1 \in v(\neg p)$ sii $0 \in v(p)$

$0 \in v(\neg p)$ sii $1 \in v(p)$

$1 \in v(p \wedge q)$ sii $1 \in v(p)$ y $1 \in v(q)$

$0 \in v(p \wedge q)$ sii $0 \in v(p)$ o $0 \in v(q)$

Dada una fórmula F con una variable libre, x (F(x/a) es el resultado de sustituir x por a en F):

$1 \in v(\forall x\mathrm{F})$ sii $1 \in v(\mathrm{F}(x/\mathrm{a}))$ para toda constante a (i.e., para todo $a \in D$, ya que $I(\mathrm{a}) = a$)

$0 \in v(\forall x\mathrm{F})$ sii $0 \in v(\mathrm{F}(x/\mathrm{a}))$ para alguna constante a (i.e., para un $a \in D$, ya que $I(\mathrm{a}) = a$)

Las conectivas '$\vee$', '$\rightarrow$', '$\leftrightarrow$' y '$\exists$', se definen a partir de '$\neg$', '$\wedge$' y '$\forall$' del modo habitual:[45]

[45] Con tres posibles evaluaciones para cada oración: {0}, {1} y {0, 1}, las tablas de verdad para las conectivas son las tablas de la lógica trivalente "fuerte" de Kleene (1952, p. 332 y siguientes), K3F para abreviar. En estas tablas (al igual que en las "débiles", K3D) las conectivas preservan los valores clásicos siempre que operan sobre oraciones que toman valores clásicos. Cuando uno de los constituyentes de una oración

$p \vee q$	sii	$\neg(\neg p \wedge \neg q)$
$p \rightarrow q$	sii	$\neg p \vee q$
$p \leftrightarrow q$	sii	$(p \rightarrow q) \wedge (q \rightarrow p)$
$\exists x$ F	sii	$\neg\forall x \neg$F

Estas equivalencias son posibles, entre otras cosas, porque las leyes de De Morgan son válidas en la lógica de Priest y derivables de las propiedades satisfechas por la negación dialetheista y la definición de '$\wedge$' dada (Priest 1999, p. 107):

(DM1) $\neg(p \wedge q)$ sii $\neg p \vee \neg q$

(DM2) $\neg(p \vee q)$ sii $\neg p \wedge \neg q$

Otro tanto se puede decir de $\neg\forall x$ F sii $\exists x \neg$F y de $\neg\exists x$ F sii $\forall x \neg$F.[46] El símbolo '$\rightarrow$' representa el *condicional material* de la lógica clásica, como veremos no es el único operador de implicación empleado por Priest.

Priest define a continuación la relación, '$\models$', de *consecuencia lógica*. Dado un conjunto de fórmulas Γ y una fórmula p,

---

compuesta mediante '$\neg$', '$\wedge$', '$\vee$', '$\rightarrow$' tiene un valor *no* clásico (i.e., {0, 1}), sistemáticamente K3D evalúa la oración compleja como {0, 1}. K3F, en cambio, no opera de este modo. Por ejemplo, si $v(p) = \{0\}$ y $v(q) = \{0, 1\}$, $v(p \wedge q) = \{0, 1\}$ en K3D, pero $v(p \wedge q) = \{0\}$ en K3F. Aunque la semántica de LP sea básicamente la de K3F, existe una diferencia importante: En K3F el único valor de verdad designado (i.e., aquel que puede adoptar la conclusión de una inferencia válida) es {1} (verdadero), en LP, los valores designados son {1} y {0, 1} (paradójico). Véase Priest 1979, III.7, pp. 227-8.

[46] Véase, por ejemplo, Priest 1998b, p. 413.

$\Gamma \models p$ si, y sólo si, para toda interpretación, <*I*, *D*>, y asignación de variables, *s*, se cumple que, si $1 \in v(q)$ para todo $q \in \Gamma$, entonces $1 \in v(p)$.

$\models p$ si, y sólo si, para toda interpretación, <*I*, *D*>, y asignación de variables, *s*, se cumple que $1 \in v(p)$.

Entre los teoremas de LP dos son muy relevantes: (T1) $\models p$ sii p es una verdad lógica en la lógica clásica de primer orden; (T2) Si $\Gamma \models p$, entonces p es una consecuencia lógica de Γ en la lógica clásica de primer orden, *pero no a la inversa* (Priest 1979, p. 230; *IC*, pp. 98, 101). T1 viene a decir que las tautologías clásicas son también tautologías en LP y viceversa. T2 afirma que las cosas que se siguen de un conjunto de fórmulas, Γ, en LP se siguen también de Γ en la lógica clásica. Sin embargo, T2 afirma también que *no* todas las cosas que se siguen de Γ en la lógica clásica se pueden establecer en LP. Una prueba de ello es que la relación de consecuencia lógica no satisface en LP las siguientes condiciones –ni las reglas de inferencia asociadas que señalamos en negrita a la izquierda– que sí satisface, en cambio, la lógica clásica:[47]

Explosión	(ECQ)	$p \wedge \neg p \models q$
Silogismo Disyuntivo	(SD)	$\neg p, p \vee q \models q$ $(p, \neg p \vee q \models q)$
Modus Ponens	(MP)	$p, p \to q \models q$
Transitividad de '→'	(TRI)	$p \to q, q \to r \models p \to r$
Reductio ad Absurdum	(RA)	$p \to (q \wedge \neg q) \models \neg p$
Antecedente Falso	(AF)	$\neg p \models p \to q$

Contraposición	(CTR)	$(p \to q) \wedge \neg q \models \neg p$

Dadas dos fórmulas, F y G, con una variable libre, x, no se cumplen:

Modus Ponens ∀	(MP∀)	$\forall x\, F, \forall x\, (F \to G) \models \forall x\, G$
Contraposición ∀	(CTR∀)	$\forall x\, (F \to G), \forall x\, \neg G \models \forall x\, \neg F$
Transitividad →/∀	(TRI∀)	$\forall x\, (F \to G), \forall x\, (G \to H) \models \forall x\, (G \to H)$

La imposibilidad de justificar estas reglas y propiedades de '$\models$' en LP hace que muchos de los resultados que el dialetheista desea establecer con respecto a la noción de verdad se compliquen. El predicado 'V' debería satisfacer PEV, que, bajo el presente marco semántico, Priest interpreta de la siguiente manera (*IC*, p. 99):

(PEV_{LP}) $\quad 1 \in v(p)$ sii $p \in I^{+}(V)$

(p es verdadera sii p es V). Además de cumplir PEV_{LP}, 'V' debería ser un predicado "exhaustivo" (véase 5.2.2.2): "Si p no es V, entonces p es falsa" (i.e., si $p \in I^{-}(V)$, entonces $0 \in v(p)$); e inconsistente, rechazando, pues: "Si p es falsa, entonces no es V" (i.e., si $0 \in v(p)$, entonces $p \in I^{-}(V)$). Ahora bien, ¿cómo hemos de interpretar '*si ..., entonces ...*' y '*sii*'? Si identificamos estas expresiones con '→' y '↔' respectivamente, no se puede asegurar la validez de PEV_{LP}, ya que el condicional material *no satisface* MP en LP y, por tanto, no podemos obtener en general 'p es V' a partir de 'p es verdadero' y de 'p es verdadero → p es V' (lo mismo ocurre al considerar la otra dirección del condicional).[48]

Priest salva este escollo argumentando que nuestra noción intuitiva de implicación *no* coincide con la de condicional material. En

---

[47] Véase, por ejemplo, Priest 1979 (III.9), pp. 228, 230. La lista es "estremecedora".

particular, la implicación que aparece en PEV_{LP} *no es* un condicional material. En el capítulo 6 de *IC*, Priest examina posibles definiciones de la noción intensional de implicación. Según Priest, la noción básica de implicación, a la que llama "entrañamiento" [*entailment*] (que representaremos mediante "=>"), debería cumplir dos requisitos (*IC*, pp. 104-5): preservar la verdad del antecedente en el consecuente (i.e., satisfacer algo parecido a MP) y preservar la falsedad del consecuente en el antecedente (i.e., satisfacer alguna forma de CTR).[49] Priest señala, sin embargo, que, como es posible tratar por separado las condiciones bajo las que un enunciado es verdadero y las condiciones bajo las que es falso,[50] también es posible pensar en una noción de implicación que *no* tenga las mismas condiciones de *falsedad* que el "entrañamiento". En particular, una implicación que *no* preserve la falsedad del consecuente en el antecedente (CTR), aunque sí MP, la verdad del antecedente en el

[48] El problema de la implicación se plantea también para el principio de abstracción (*IC*, p. 102).

[49] Priest (1979 IV.5, p. 233; *IC* 6.2, pp. 103-4) señala un requisito adicional. Para que la noción de implicación definida *no sea trivial*, debe excluir los principios de "*absorción*": $p \to (p \to q) \vdash p \to q$, y de "*aserción*": $(p \wedge (p \to q)) \to (p \to q)$. Estos principios, combinados con MP, permiten inferir cualquier oración, como muestra la paradoja de Curry [17] (véase 2.8.1).

[50] Como vimos en 5.2.2.2, para un dialetheista los hechos positivos (aunque no los negativos) a favor de F($\underline{p}$) son *independientes* de los hechos positivos a favor de V($\underline{p}$). Desde esta perspectiva, la falsedad de una oración (basada en hechos positivos) es independiente de su verdad. Esta independencia parcial de las nociones de verdad y falsedad es el resultado de afirmar que no son propiedades excluyentes. La verdad de una oración no elimina su falsedad y viceversa: V($\underline{p}$) y F($\underline{p}$) se pueden establecer por separado. No obstante, verdad y falsedad no son del todo independientes. Al ser propiedades exhaustivas algunas veces la única razón que tenemos para afirmar F($\underline{p}$) es (el hecho "negativo" de) que no podemos establecer V($\underline{p}$). Esto justifica que, junto al predicado 'V', tengamos también 'F', un predicado parcialmente independiente que cumple las equivalencias F($\underline{p}$) sii $\neg p$ sii V($\underline{\neg p}$).

consecuente (*IC*, pp. 109-10). Representaremos el "entrañamiento" mediante "=>" y la conectiva que no respeta CTR mediante "$\Rightarrow$".[51] Según Priest, la conectiva que podemos encontrar en PEV_{LP} es como "$\Rightarrow$", ya que respeta MP y permite, por tanto: $p, p \Rightarrow q \models q$, pero no respeta CTR, de modo que: $p \Rightarrow q, \neg q \models \neg p$, es inválido. El rechazo de CTR y "=>" (y la consiguiente elección de "$\Rightarrow$") en PEV_{LP} responde al siguiente motivo: Si la implicación de PEV_{LP} satisficiese CTR, podríamos demostrar que 'V' es consistente, i.e., podríamos demostrar "Si p es falso, entonces p no es V" (*IC*, p. 100).[52]

5.2.3.2 LP es mínimamente inconsistente: razonamientos "cuasi-válidos"

Pese a la introducción de nuevas conectivas intensionales que puedan otorgar validez general a las inferencias que el dialetheista desea efectuar mediante PEV, las restricciones que la lógica paraconsistente de Priest introduce en nuestras prácticas demostrativas son alarmantes.

[51] Priest ofrece tentativamente más de una conectiva de tipo "$\Rightarrow$" (*IC*, pp. 109-10) y da indicaciones con respecto al *entailment*, "=>", que pueden concretarse en más de una conectiva (*IC*, pp. 104-8). Sí deja claro, no obstante, que su noción de *entailment* no coincide con la de implicación estricta o con la noción de *entailment* de la lógica de la relevancia (*IC*, pp. 111-4).

[52] Esto es, apelando a PEV, CTR y la definición de 'falso' ('F') podemos demostrar: $F(\underline{p}) => \neg V(\underline{p})$.

1. $F(\underline{p})$ (suposición)
2. $\neg p$ (definición de 'F' + PEV [$F(\underline{p}) <=> V(\underline{\neg p}) <=> \neg p$]: 1)
3. $V(\underline{p}) => p$ (PEV)
4. $\neg V(\underline{p})$ (CTR: 2, 3)
5. $F(\underline{p}) => \neg V(\underline{p})$ (Introducción '=>': 1-4)

(He utilizado conectivas del tipo '=>' y '<=>' porque son las que satisfacen CTR. Vemos así que, si '=>' fuese el condicional que aparece en PEV_{LP}, podríamos demostrar que F y V son *excluyentes*.)

Priest es consciente de ello y busca minimizar el impacto de este hecho haciendo hincapié en dos ideas:[53] (1) LP se puede presentar como un cálculo *mínimamente inconsistente*; y (2) *todas* las reglas *clásicas* de inferencia son *válidas* en LP siempre y cuando las conectivas operen sobre oraciones que *no sean dialetheias*.

De acuerdo con (2), *sólo* las dialetheias hacen fracasar la lista de reglas expuesta en 5.2.3.1. Cuando las oraciones relevantes tienen valores de verdad clásicos, MP, SD, CTR, ECQ, etc. pasan a ser reglas de inferencia fiables. Priest llama "cuasi-válidas" a las reglas que preservan la verdad de las premisas en la conclusión siempre que las premisas reciben valores *clásicos*.[54] Por desgracia, no existe un modo de identificar sintácticamente las dialetheias, lo que nos permitiría "señalarlas" excluirlas del alcance de MP, SD, etc. y reformular estas reglas de modo que pasen a ser leyes completamente generales. Priest (*IC*, pp. 138-40) ilustra este hecho mediante una estrategia fallida (para indicar sintácticamente que *p no es* una dialetheia) consistente en añadir a un razonamiento la premisa: $\neg(p \wedge \neg p)$. Esta estrategia falla porque nada nos garantiza que $\neg(p \wedge \neg p)$ y $p \wedge \neg p$ sean ambas verdaderas (i.e., dialetheias) y, por tanto, que p también lo sea. Después de todo, vimos en 5.2.2.3 y 5.2.3.1 que, pese a ser inconsistente, LP contiene entre sus

---

[53] Estas ideas aparecen en *IC*, capítulo 8, también en Priest 1979, IV, pp. 231-7 y en Priest 1991.

[54] En este sentido y con respecto al impacto que supone sacrificar las reglas enumeradas en 5.2.3.1, Priest (1979, p. 231) dice "We can soften the blow to the intuitions a little by pointing out that although these inferences are not generally valid, they are valid provided all the truth values involved are classical (i.e., true only or false only). Let us call a rule that is truth preserving only under such conditions 'quasi-valid'."

verdades lógicas a $\models \neg(p \wedge \neg p)$ (para toda oración $p \in$ LP, aun cuando p sea una dialetheia).[55]

El uso de reglas "cuasi-válidas" es, no obstante, perfectamente racional y está plenamente justificado. La probabilidad estadística de tropezar con una dialetheia es nimia, próxima a cero (aunque no igual a cero según Priest). Por este motivo, leyes cuasi-válidas como SD o MP permiten en la práctica hacer inferencias que preservan la verdad de las premisas casi en el cien por cien de los casos, sólo cuando tropezamos con paradojas lógicas su efectividad se suspende (*IC*, pp. 144). La baja probabilidad de que una contradicción sea una dialetheia hace que sea *racional*, incluso para un dialetheista, *rechazar* una contradicción. De hecho, el lógico dialetheista asume por defecto (lo cual es razonable) que las oraciones con que opera habitualmente son consistentes y aplica, por tanto, reglas cuasi-válidas. Sólo cuando tiene razones independientes de peso para creer en la verdad de una contradicción suspende la validez de su hipótesis de partida (i.e., la consistencia de las oraciones con que opera) y rechaza el uso de reglas cuasi-válidas (*IC*, p. 145).[56] El objetivo de un dialetheista es, pues, minimizar, no sólo el número de

---

[55] '*p es* una dialetheia' podría ser una dialetheia, lo que haría verdadera también a '*p no es* una dialetheia'. Según Priest, "if a situation is inconsistent, adding an extra premise to the effect that it is consistent will not change the situation, but merely multiply the inconsistencies. There is no statement that can be made which *forces* α to behave consistently. This is one of the hard facts of dialetheic life." *IC*, p. 140.

[56] Priest resume en una "máxima metodológica" este modo de proceder: "Unless we have especific grounds for believing that the crucial contradictions in a piece of quasi-valid reasoning are dialetheias, we may accept the reasoning." *IC*, p. 145.

inconsistencias que ha de contener nuestro lenguaje, sino también el impacto que éstas tienen en nuestras prácticas inferenciales cotidianas.[57]

5.2.3.3 Modelos inconsistentes

Como es obvio, Priest se preocupa por el estudio de modelos inconsistentes *no triviales*. Una de las mejores maneras de mostrar que la idea de "dialetheia" es coherente, consiste en construir modelos que satisfagan ciertas contradicciones, pero no todas. Ésta es una de las principales líneas de investigación de Priest (*BLT*, p. 170). En este sentido, uno de sus intereses prioritarios, en consonancia con lo dicho en el apartado anterior, es aclarar el significado de "*mínimamente inconsistente*" ("*mi*" en lo sucesivo) al hablar de la aplicación de LP a un modelo inconsistente. El objetivo final es mostrar que un dialetheista puede apropiarse del razonamiento clásico suspendiendo la validez de las leyes clásicas sólo en casos extraordinarios (i.e., en presencia de dialetheias). Priest (1991, p. 323) llama a la lógica *mi* resultante LPm. El primer paso consiste en definir una relación de orden, '$<$', que permita comparar modelos inconsistentes (ibíd., p. 325). Priest identifica el núcleo inconsistente de un modelo $M = <I, D>$ con el conjunto, $M! \subseteq D$,

[57] Priest (1991, pp. 321-2) afirma que las lógicas paraconsistentes sólo tienen dos rutas alternativas para incorporar reglas como SD a las que resulta difícil renunciar. Una, la más común (pero también la menos exitosa), consiste en relacionar los fallos de razonamientos en los que se aplica SD con la presunción de que el condicional material es el condicional "genuino". La estrategia consiste, pues, en dar una formulación alternativa de condicional. La segunda estrategia es la que elige Priest aquí (aunque también emplea la primera, como vimos en 5.2.3.1): incorporar las reglas clásicas a nuestras prácticas demostrativas describiéndolas como "cuasi-válidas" y restringiéndolas a oraciones consistentes.

de oraciones *atómicas* cuya evaluación es inconsistente: $M! = \{ p \in D: p$ es atómica y $v(p) = \{0, 1\}\}$.[58]

A partir del "tamaño" de $M!$, se establece una gradación de modelos, M, de más a menos inconsistentes. Dados dos modelos $M = <I, D>$ y $M' = <I', D'>$ con idéntico dominio ($D = D'$), M' es *más inconsistente* que M si, y sólo si, $M'!$ contiene las dialetheias atómicas de M y algunas más, i.e., $M < M'$ sii $M! \subset M'!$. El modelo más inconsistente de todos es el *trivial*, en el que todas las proposiciones atómicas son dialetheias y, por ende, *todas* las oraciones (ibíd., lema p. 324). Los modelos menos inconsistentes son los clásicos, donde $M! = \varnothing$. Una vez definida la relación de orden, podemos caracterizar las nociones de (I) "*modelo mi*" y (II) "*consecuencia mi*" como sigue (ibíd., p. 325):

[58] Una dialetheia se produce cuando $1 \in v(p \wedge \neg p)$, ahora bien, esto sólo es posible si $v(p) = \{0, 1\}$. Lo que determina la parte inconsistente de un modelo es el conjunto de oraciones simples que reciben la misma evaluación que p, i.e., $\{0, 1\}$. Priest (1991, p. 325) llama a estas oraciones "*atómicas*" porque no parecen contener símbolos lógicos. Sin embargo, esto es engañoso, ya que la mayoría de dialetheias básicas son, como hemos visto en los capítulos 2 y 3, *contradicciones* de inclusión y entrañan, además, algún tipo de autorreferencia. Esto quiere decir que las dialetheias básicas *no son* oraciones atómicas, son oraciones que contienen, *al menos*, el símbolo de *negación* y que (habitualmente por ser autorreferentes) satisfacen la equivalencia $p \leftrightarrow \neg p$. A menos que tratemos la negación como un símbolo no lógico, no parece correcto hablar de dialetheias como oraciones atómicas. Es verdad que la oración '$V(\underline{\mu})$' (donde $\underline{\mu}$ = '$\neg V(\underline{\mu})$') es *atómica* y es una dialetheia, pero los problemas que genera '$V(\underline{\mu})$' se derivan en primera instancia de μ (el mentiroso), que *no es* atómica. Desde este punto de vista, apelar a oraciones atómicas para cuantificar la inconsistencia de una teoría, aunque parece la maniobra más razonable, no está nada claro. Quizá la distinción entre hechos negativos y positivos a la hora de establecer la verdad de '$\neg p$' tenga relevancia en este sentido. Tal vez una negación, '$\neg p$', que sea verdadera debido a hechos *positivos* (independientes de cualquier hecho que avale la verdad de 'p') pueda considerarse una oración que refleja hechos *atómicos*. Pero todo esto parece bastante enrevesado, tortuoso y poco o nada convincente.

(I) M es un *modelo mi* de un conjunto de oraciones Γ ($M \models_m \Gamma$) si, y sólo si, (1) M es un modelo de Γ; y (2) cualquier modelo *menos* inconsistente que M, *no es* un modelo de Γ:

$M \models_m \Gamma$ sii (1) $M \models \Gamma$ y (2) si $M' < M$, entonces *no es el caso* que $M' \models \Gamma$

(II) p es una *consecuencia mi* de Γ si, y sólo si, todo modelo *mi* de Γ es un modelo de p:

$\Gamma \models_m p$ sii para todo M, si $M \models_m \Gamma$, entonces $M \models p$.

La lógica LPm opera siempre con modelos y consecuencias *mi* de un conjunto Γ de oraciones. Si Γ *no* incluye una contradicción (o fórmulas de las que se siga una contradicción –por ejemplo: α, $\neg\alpha$–), entonces los modelos *mi* de Γ *serán clásicos* y, por tanto, las reglas cuasi-válidas de inferencia (i.e., las clásicas) podrán usarse para extraer consecuencias de Γ mediante LPm. Si Γ incluye, por contra, (fórmulas de las que se sigue) una contradicción: $p \wedge \neg p$ ($p!$, para abreviar), entonces los modelos *mi* de Γ *serán inconsistentes*, ya que sólo un modelo inconsistente permite $1 \in v(p!)$. En este caso LPm suspenderá el uso de reglas cuasi-válidas. Esto implica, entre otras cosas, que LPm *no es* un cálculo *monotónico*. Consideremos un ejemplo, supongamos que $\Gamma = \{p, \neg p \vee q\}$, en este caso tenemos (por II): $\Gamma \models_m q$, ya que cualquier modelo *mi* que satisfaga todas las fórmulas de Γ *es clásico* (por I) y satisface q (por SD). Ahora bien, si consideramos $\Gamma^* = \Gamma \cup \{p!\}$, entonces *no es el caso* que $\Gamma^* \models_m q$, ya que, para que un modelo *mi* satisfaga *todas* las fórmulas de Γ^* (como reza I), ha de ser *inconsistente* y contener $1 \in v(p!)$, lo que impide que q sea una consecuencia *mi* de Γ^* por SD. Hay modelos *mi* de Γ^* tales que $1 \notin v(q)$ (q sólo será verdadera en un modelo *mi* de Γ^* que haga a q verdadera *con independencia* de los valores que asigne a p). De este modo, añadir premisas a un conjunto de oraciones puede bloquear algunas de las inferencias (cuasi-válidas) que podíamos hacer antes de añadir dichas premisas. Además de no monotónica, LPm satisface también otras propiedades: Sean Γ^{LP}, Γ^{LPm} y Γ^{CL} las consecuencias que uno puede sacar de Γ en LP, LPm y la lógica clásica respectivamente, se cumple entonces que $\Gamma^{LP} \subseteq \Gamma^{LPm} \subseteq \Gamma^{CL}$ (en general, la inclusión es *propia*). Además, si Γ es consistente en términos clásicos, $\Gamma^{LPm} = \Gamma^{CL}$. Todos estos hechos muestran, según Priest (ibíd., p. 326), que LPm hace efectiva la presunción por defecto de la consistencia de un conjunto, Γ, de oraciones.

LPm permite, en general, realizar más inferencias que LP a partir de Γ ($\Gamma^{LP} \subseteq \Gamma^{LPm}$), esto podría plantear un problema que Priest también aborda. Es un hecho que LPm puede demostrar *más* contradicciones que LP (ibíd., pp. 326-7), la cuestión es ¿podría Γ^{LPm} desembocar en el trivialismo, a diferencia de Γ^{LP}? La respuesta es *no*. Priest demuestra que cuando las consecuencias que uno puede extraer de Γ mediante LP (Γ^{LP}) no son triviales, *tampoco* lo son las que se pueden extraer de Γ mediante LPm (Γ^{LPm}). Priest llama a esta propiedad de LPm "tranquilizadora" [*reassurance*] y demuestra su cumplimiento en lógica proposicional y de primer orden (ibíd., pp. 327-30).[59]

Más fundamental aún que la pregunta por el significado de "modelo *mi*" es la pregunta acerca de *cómo* construir un *modelo inconsistente* para un lenguaje L sobre el que aplicar LP. Priest lleva a cabo esta tarea en diferentes escritos.[60] La idea básica es buscar un modelo consistente, *M*, y crear a partir de él uno nuevo, *M'*, que "colapse", "contraiga" o "fusione" varios elementos del dominio de *M* en un único elemento, *e*, del dominio de *M'*. Se trata de conseguir que *e* conserve *todas* las propiedades que tenían por separado los miembros de

[59] Este entramado permite construir un cálculo formal que otorgue sentido a la idea de que, "en ausencia de contradicciones, podemos presumir que las oraciones con las que operamos son consistentes y aplicar reglas clásicas de inferencia". Sin embargo, Priest no nos dice en ningún momento *cómo* determinar si una contradicción es verdadera (i.e., una dialetheia aceptable) o sólo falsa y, por tanto, rechazable. Las contradicciones rechazables *no deberían forzarnos* a suspender el uso de reglas clásicas, sin embargo, LPm prima siempre la "presunción de inocencia", por así decirlo. Cuando una contradicción aparece, se suspende el uso de reglas clásicas *por si fuese verdadera* (todo modelo *mi* de un conjunto de oraciones Γ que contenga una contradicción *ha de ser inconsistente*). Esto choca frontalmente con el hecho de que *la inmensa mayoría* de contradicciones *no son verdaderas*. La presunción de culpabilidad es, pues, más plausible. LPm nos fuerza, además, a redefinir algunas reglas cuasi-válidas de inferencia que operan con contradicciones, de lo contrario, *nunca se aplicarían*. En lógica clásica, por ejemplo, definimos '$\perp$' del siguiente modo: $B \wedge \neg B \leftrightarrow \perp$. Esto nos permite definir RA como: [A] ... $B \wedge \neg B \Rightarrow \neg A$ ("si llegamos *a una contradicción* suponiendo A, negamos A"). LPm debe reformular '$\perp$' para poder utilizar RA (el candidato más plausible es '$\forall x\ V(\underline{x}) \leftrightarrow \perp$', i.e., absurdo significa "todo es verdadero", RA se convierte en: [A] ... $\forall x\ V(\underline{x}) \Rightarrow \neg A$). Pese a todo, el problema de fondo sigue intacto: Una de las mayores taras del dialetheismo es la *indecidibilidad* de las dialetheias. Volveremos a este tema más adelante (5.3.1-2) por su relación con el uso que hacemos de las contradicciones mediante RA para *rechazar* creencias.

[60] Una presentación resumida de todas estas cuestiones se puede encontrar en *BLT*, pp. 170-5. Véase también Priest 1991 o Priest 1994b, pp. 137-8, 146-7.

M "fusionados" en e, desarrollando así propiedades contradictorias.[61] Veamos con más detalle cómo se construye un modelo inconsistente.

Partiendo de un modelo consistente $M = <I, D>$ cuyo dominio, D, tiene cardinalidad κ, construimos un modelo inconsistente $M^{\sim} = <I^{\sim}, D^{\sim}>$ con dominio, $D^{\sim}$, de cardinalidad λ inferior a κ (λ < κ). Para ello, definimos una relación de equivalencia, "∼", sobre los elementos de D y extraemos las *clases de equivalencia* que "∼" determina. Las clases de equivalencia así obtenidas serán los *miembros* de $D^{\sim}$, que se convierte en el resultado de "fusionar" o "contraer" varios elementos de D en un número inferior de elementos (siempre y cuando haya *menos* clases de equivalencia en $D^{\sim}$ que elementos contiene D).

Ejemplo: Sea $D = \{a, b, c, d\}$ y "∼" una relación de equivalencia (i.e., reflexiva, simétrica y transitiva) tal que: $a \sim b$ y $c \sim d$ ("∼" determina las siguientes clases de equivalencia: $[a] = \{a, b\}$ y $[c] = \{c, d\}$).[62] A partir de D y "∼", definimos ahora: $D^{\sim} = \{[a], [c]\}$. Como se puede apreciar, D tiene cardinalidad 4, mientras que $D^{\sim}$ tiene cardinalidad 2; $D^{\sim}$ es el resultado de fusionar varios miembros de D en uno sólo, a saber, una clase de equivalencia (a y b se contraen en $[a] \in D^{\sim}$; c y d se contraen en $[c] \in D^{\sim}$).

[61] Este tipo de técnicas para producir modelos inconsistentes (incluido el "*Collapsing Lemma*" que veremos más adelante) es de uso común en lógicas paraconsistentes, no sólo las emplea Priest. R. K. Meyer es su iniciador en los 70 (véase *BLT*, p. 170).

[62] "∼" no puede ser la relación de identidad, "=", u otra equivalente, ya que se trata de *contraer* o *reducir* los elementos de D mediante "∼", cosa que "=" no consigue ("=" asigna cada elemento de D a sí mismo). Para asegurar este requisito, debe haber dos elementos de D tales que $a \neq b$ y $a \sim b$.

Una vez definido el dominio $D^{\sim}$ de $M^{\sim}$, podemos definir $I^{\sim}$ (su función de interpretación con respecto al lenguaje L) del siguiente modo: Si c es una constante de L que denota a c en M, entonces c denotará a $[c]$ en $M^{\sim}$; si P es un predicado n-ádico de L, entonces $<[c]_0, ..., [c]_{n-1}>$ forma parte de la extensión positiva (negativa) de P en $M^{\sim}$ si, y sólo si, $\exists x_0 \in [c]_0, ..., \exists x_{n-1} \in [c]_{n-1}$ tales que $<x_0, ... , x_{n-1}>$ forma parte de la extensión positiva (negativa) de P en M. La definición de $I^{\sim}$ completa la definición de $M^{\sim}$. Para ver que $M^{\sim}$ es un modelo inconsistente recurrimos al siguiente resultado formal:

Collapsing Lemma. Sea φ una fórmula cualquiera de L y sea v igual a 1 o a 0. Se cumple entonces que si v $\in$ $\nu(\varphi)$ en el modelo M, entonces v $\in$ $\nu(\varphi)$ en $M^{\sim}$ (i.e., $\nu(\varphi) \subseteq \nu^{\sim}(\varphi)$)[63]

Lo que este lema afirma es que, cuando contraemos o "colapsamos" M en $M^{\sim}$, las fórmulas de L *no pierden* valores de verdad, sólo pueden *ganarlos*. Una fórmula que es verdadera (falsa) en M no dejará de serlo en $M^{\sim}$, pero podría pasar a ser, *además*, falsa (verdadera), generando así una dialetheia.

---

[63] *BLT*, p. 172. Formulaciones equivalentes de este lema son: "para toda fórmula φ, $\nu(\varphi) \subseteq \nu^{\sim}(\varphi)$". (Priest 1991, p. 329); o, simplemente, "para toda fórmula φ, si φ es verdadera/falsa en M, entonces φ es verdadera/falsa en $M^{\sim}$" (Priest 1994b, p. 346). Una demostración del lema (por recursión sobre la estructura de las fórmulas de L) se puede encontrar en Priest 1991, p. 329. Priest lo identifica con el teorema descendente de Löwenheim-Skolem: "The Collapsing Lemma is the ultimate downward Löwenheim-Skolem theorem." (*BLT*, p. 1972; también, con respecto a modelos de la aritmética, en Priest 1994b, p. 339).

Ejemplo: Tomemos dos fórmulas de L: P(a) y P(b), que sean respectivamente verdadera y falsa en M (v(P(a)) = {1} y v(P(b)) = {0}). La interpretación de a y b en M es: I(a) = a, I(b) = b. Construyamos ahora un modelo $M^{\sim}$ tal que $a \sim b$, se cumple entonces que: $a, b \in [a]$. La interpretación de a, b en $M^{\sim}$ es, pues: $I^{\sim}$(a) = $I^{\sim}$(b) = [a]. La pregunta ahora es, ¿qué evaluación reciben P(a) y P(b) en $M^{\sim}$? La respuesta es P(a) y P(b) *son dialetheias* en $M^{\sim}$: $v^{\sim}$(P(a)) = $v^{\sim}$(P(b)) = {0, 1}.

Esto se debe a que P(a) expresaba *un* hecho verdadero en M y P(b) *un* hecho falso, sin embargo, en $M^{\sim}$ ambas oraciones expresan el *mismo* hecho, ya que hemos fusionado a y b (los referentes en M de a y b respectivamente) para obtener [a] (el referente de a y b en $M^{\sim}$). Por otra parte, las definiciones de $D^{\sim}$ e $I^{\sim}$ garantizan que las propiedades que en M tenían por separado los elementos que integran en $M^{\sim}$ la clase de equivalencia [c] *se acumulen* en [c] (éste es el hecho sobre el que gira la demostración del "*Collapsing lemma*"). Se cumple así: v(P(a)) = $\{1\} \subseteq \{0, 1\}$ = $v^{\sim}$(P(a)) y v(P(b)) = $\{0\} \subseteq \{0, 1\}$ = $v^{\sim}$(P(b)), como reza el lema que, en general, garantiza $v(\varphi) \subseteq v^{\sim}(\varphi)$ (la preservación en $M^{\sim}$ de los valores de verdad que cualquier fórmula, φ, de L toma en M).

Priest ofrece modelos inconsistentes para entidades que encajan en el Esquema de Inclusión (*BLT*, pp. 173-5), pero no se detiene ahí. Como vimos en 5.2.1.1 al hablar del teorema de Gödel, Priest también considera que la aritmética es inconsistente, así pues, también construye modelos inconsistentes para el lenguaje de la aritmética, L_A.[64] Junto al teorema de Gödel, Priest (1994b) ofrece nuevos argumentos a favor de la inconsistencia de la aritmética y de sus modelos interpretativos. El primero está basado en el argumento de la indeterminación semántica de Putnam. Un argumento en contra de los modelos inconsistentes de la aritmética consistiría en señalar simplemente que, N (el modelo consistente estándar), *es el correcto*. Priest cuestiona esta afirmación. Para empezar (y olvidándonos de los modelos inconsistentes por un momento), es un hecho que la aritmética tiene un sinfín de modelos *consistentes* diferentes (y no isomorfos, aunque equivalentes). Priest

---

[64] En 1994b (pp. 337-8, 346-7), Priest toma como modelo consistente de L_A el modelo estándar de la aritmética, N, y construye a partir de él modelos inconsistentes *finitos*, N^-. El procedimiento es el ya descrito. El dominio de partida es esta vez el conjunto de los naturales: Elegimos un número natural n y definimos la relación de equivalencia "$\sim$"del siguiente modo: Si $m < n$, entonces $m \sim x \leftrightarrow m = x$ (i.e., todos los números *menores* que n son *sólo* equivalentes a sí mismos); si $m \geq n$, entonces $m \sim n$ (i.e., n y todos los números *mayores* que n son equivalentes a n). De este modo formamos n + 1 clases de equivalencia, las primeras n clases ($[0] = \{0\}$, $[1] = \{1\}$, ..., $[n - 1] = \{n - 1\}$) se corresponden con los primeros n números naturales; la clase n + 1 ($[n] = \{n, n + 1, n + 2, ...\}$) *aglutina* o *fusiona* al *resto* de números naturales *en uno solo*: n. Si representamos los *numerales* mediante un subrayado, el lenguaje de la aritmética se interpreta en N^- de tal forma que $\underline{0}$ representa a [0]; $\underline{1}$ a [1]; y así sucesivamente hasta llegar a $\underline{n}$ que representa a $[n] = \{n, n + 1, n + 2, ...\}$. Todas las propiedades de n y de los números mayores que n se "colapsan" en N^- generando así contradicciones como $n = n + 1$ y $n \neq n + 1$; $n = n$ y $n \neq n$ (ya que, pese a ser idénticos en N^-, n y n+ 1 conservan en N^- las mismas propiedades que tenían en N, por ejemplo: $n = n$, $n + 1 = n + 1$ y $n \neq n + 1$). El "*Collapsing Lemma*" garantiza que $v^-(\underline{n} = \underline{n} + \underline{1}) = \{0, 1\}$. Aunque podemos construir tantos modelos inconsistentes de la aritmética como números naturales, n, diferentes elijamos a tal efecto, Priest sugiere que el n elegido sea un número lo suficientemente grande como para preservar nuestras intuiciones con respecto a la consistencia de las operaciones aritméticas que puede efectuar o imaginar un ser humano: "Let us, henceforth, fix *n* as some incredibly large number, say a number larger than the number of combinations of fundamental particles in the cosmos, larger than any number that could be sensibly specified in a lifetime, so large that it has no physical meaning or psychological reality." (Priest 1994b, p. 338).

(1994b, pp. 338-40) se pregunta qué hace de N un modelo de L_A mejor que cualquier otro modelo consistente y rechaza en este sentido una serie de argumentos gödelianos que apelan a facultades mentales especiales que nos permitirían "intuir" de forma directa ciertas verdades matemáticas sobre números. Para Priest, de existir estas facultades (cosa que duda), no nos permitirían decidir la cuestión de cuál es la correcta interpretación (consistente) de L_A. Así pues, Priest apela al argumento de la indeterminación semántica de Putnam (cuya versión general para lenguajes naturales no suscribe) para afirmar que existen diversos modelos consistentes de L_A, pero que carecemos de criterios para privilegiar uno sobre el resto. Nada determina que N sea la interpretación "correcta" de L_A. Una vez "derogados" los privilegios de N, Priest ofrece argumentos a favor de usar modelos inconsistentes en la interpretación de L_A. El primero (ibíd. pp. 340-41) hace referencia a la indiscernibilidad en la práctica de las operaciones que llevan a cabo el matemático paraconsistente y el consistente, dada la elección de modelo inconsistente hecha –véase la cita final de la nota anterior sobre la elección de un número natural n para construir un modelo inconsistente finito de L_A–.[65] El segundo argumento está relacionado, una vez más, con la racionalidad de aceptar ciertas contradicciones. Se insiste en el teorema de Gödel como paradoja (ibíd. pp. 343-4) y se ofrece una nueva contradicción de inclusión: "el argumento de Petersen" (quien no comparte la

[65] Esta afirmación es, cuando menos, polémica. La práctica habitual de la aritmética involucra afirmaciones generales sobre todos los números naturales y está claro que alguien que opera con modelos finitos no hace el mismo tipo de afirmaciones que

interpretación que Priest hace de su argumento), consistente en una demostración de la existencia de un número natural, x, tal que $x = x + 1$.[66] Por último Priest señala las ventajas que tiene la aritmética inconsistente frente a la consistente repasando puntos concretos del programa de Hilbert (decidibilidad, completud, etc.) que podría satisfacer una aritmética inconsistente, a diferencia de una consistente (ibíd., p. 345). Los estudios formales más completos de Priest con respecto a modelos inconsistentes de la aritmética (finitos y no finitos) se pueden encontrar en Priest 1997b y 2000c.

Para concluir este apartado, añadiré que es posible utilizar los modelos kripkeanos de punto fijo para construir un modelo inconsistente no trivial que respete la semántica de LP. Dowden (1984) es el primero en señalar este hecho. Tomemos la interpretación de un lenguaje L que determina el punto fijo *mínimo* de la función de salto, *s*, que definimos en la jerarquía de interpretaciones Kripkeana (véase 4.5.2). En esa interpretación, a la que llamaremos *K*, todas las oraciones que no contienen el predicado 'verdadero' ('V') reciben un valor de verdad clásico. Entre las oraciones que contienen a 'V' algunas son

---

alguien que opera con modelos infinitos (aunque en Priest 2000c se consideran modelos infinitos inconsistentes de la aritmética).

[66] Ibíd, pp. 341-2: Sea π una *abreviación* de "El menor número al que se refiere esta descripción (ó 0 si no tiene referencia) + 1". π debe tener referencia, ya que si no la tuviese, su referente sería, por definición, 1 (i.e., 0 + 1). Así pues: (*) π = (el menor x tal que 'π' refiere a x) + 1. Obviamente, 'π' se refiere a π y, como la referencia es única, π es el menor x al que se refiere 'π', ahora bien, por (*), $\pi = \pi + 1$, de donde se sigue $\exists x\,(x = x + 1)$. (Esta contradicción de inclusión es en el fondo una versión de la paradoja de Barwise y Moss [12] (véase 2.10.1 en el apéndice del capítulo 2). La diferencia más relevante es la afirmación de la existencia de π, pero ambas coinciden plenamente si construimos un argumento a favor de la existencia de τ en [12].)

fundamentadas y otras no. En K todas las oraciones fundamentadas reciben un valor de verdad clásico y las no fundamentadas (entre las que se cuentan las paradojas de inclusión) carecen de valor de verdad. Obviamente, el modelo K no satisface la semántica de LP, sin embargo, K se puede modificar para obtener un modelo inconsistente J que cumpla este objetivo. Dowden (ibíd., pp. 127-9) toma la evaluación que reciben las oraciones que no contienen a 'V' en K y *la mantiene fija* en J: Si la oración p no contiene a 'V', entonces $J \models p$ sii $K \models p$ (una oración, p, es falsa en un modelo M sii $\neg p$ es verdadera en M, así pues, p es verdadera en K sii $J \models p$; p es falsa en K sii $J \models \neg p$). A continuación, Dowden modifica la interpretación de 'V' en K del siguiente modo: La *extensión* de V en J (i.e., V^{+}_{J}) es el conjunto de oraciones que no son declaradas falsas por K: $V^{+}_{J} = \{p$: *no* es el caso que $K \models \neg p\}$; la *antiextensión* de V en J (V^{-}_{J}) es el conjunto de oraciones no declaradas verdaderas en K: $V^{+}_{J} = \{p$: *no* es el caso que $K \models p\}$. Dowden obtiene los siguientes resultados con respecto a J: (1) Todas las oraciones cumplen: $J \models V(\underline{p})$ sii $J \models p$ y $J \models \neg V(\underline{p})$ sii $J \models \neg p$ (el subrayado se utiliza para construir nombres de términos lingüísticos). (2) Todas las oraciones reciben un valor de verdad: $\forall p\ J \models p$ o $J \models \neg p$. (3) Algunas oraciones reciben dos valores de verdad: $\exists p\ (J \models p\ \&\ J \models \neg p)$.[67] (4) No todas las oraciones reciben dos

[67] La definición que Dowden ofrece de J es insatisfactoria en un punto: J hace verdaderas a todas las oraciones que *no* son falsas en K y falsas a todas las que *no* son verdaderas. Esto significa que cuando una oración *carece* de valor de verdad en K, será verdadera en J (por no ser falsa en K) y falsa (por no ser verdadera en K). El problema radica en que, en K, *todas* las oraciones no fundamentadas *carecen* de valor de verdad, tanto las contradicciones de inclusión como la oración del veraz ($\underline{v}$ = 'V($\underline{v}$)'). Ahora bien, *sólo* las primeras son dialetheias, $\underline{v}$ es meramente falsa para Priest. Una manera de

valores de verdad (por ejemplo, las fundamentadas de *K*): $\exists p \neg(J \models p \,\&\, J \models \neg p)$.

5.3 Críticas al dialetheismo

Hasta aquí hemos examinado la defensa que Priest hace del dialetheismo, a partir de ahora nos centraremos en las críticas de sus detractores, algunas de ellas se han insinuado ya a lo largo de nuestra exposición, en la presente sección haremos hincapié en ellas. En contra de lo que muchos asumen, el dialetheismo es una posición seria y respetable, como hemos tenido ocasión de comprobar. Aunque algunas de las críticas que abordaremos a continuación plantean problemas muy serios para el dialetheista, nuestra preocupación aquí no será tanto establecer si el dialetheismo está o no en lo cierto, como establecer si, en su contienda con otras teorías rivales por ser "la teoría correcta", el dialetheismo tiene una ventaja clara, sobre todo –dado que esa es su motivación original– a la hora de ofrecer una solución satisfactoria a las paradojas de inclusión.

La primera cuestión que se nos plantea, pues, es en qué sentido el dialetheismo puede ofrecer una solución *más satisfactoria* a las paradojas reflexivas que sus rivales: Según Priest, (0) la fuerza de las paradojas de

---

subsanar este problema es tener en cuenta otros puntos fijos de la jerarquía kripkeana, *además* de *K*. Podemos definir un modelo J^* igual a J en todo excepto en un aspecto: J^* declara falsas a las mismas oraciones que J (las no verdaderas en *K*), pero declara verdaderas *sólo* a las oraciones verdaderas en *K* (las verdades fundamentadas) y a las oraciones que carecen de valor de verdad en *cualquier* punto fijo kripkeano (no sólo en *K*). Éstas últimas son las contradicciones de inclusión, ya que oraciones como el veraz son verdaderas o falsas en algunos puntos fijos de la jerarquía kripkeana.

inclusión reside en que esgrimen argumentos sólidos basados en principios generales intuitivamente correctos para aceptar tanto la verdad de α como la de $\neg\alpha$. La solución de Priest sería más satisfactoria que la de sus rivales en la medida en que hace justicia a dichos argumentos aceptando la conclusión que nos imponen. Al rechazar la verdad de α y $\neg\alpha$, los oponentes del dialetheismo han de rechazar la validez de alguno de los argumentos que sustentan dicha contradicción, para Priest las consecuencias de esta postura son insatisfactorias esencialmente por cuatro motivos: (1) Por bloquear la generalidad de ciertos principios básicos del razonamiento (fundamentalmente el esquema tarskiano, PEV; y el de abstracción, PEA) mediante distinciones entre lenguaje y metalenguaje, o teoría y metateoría, que desembocan en restricciones expresivas y mutilaciones de nuestras prácticas inferenciales. (2) Por ofrecer razones *ad hoc* (no independientes del deseo de evitar contradicciones) para restringir PEV o PEA. (3) Por no ofrecer una solución general a las paradojas de cierto campo, sino teorías y lenguajes incompletos que sólo pueden resolver coherentemente las paradojas de inclusión que se producen en las teorías o lenguajes que toman por objeto, pero no las que se producen en su propio seno. (4) Por no ofrecer una solución uniforme a problemas que comparten la misma estructura.

Si damos un giro dialéctico y adoptamos ahora la perspectiva de los detractores del dialetheismo, nos daremos cuenta, sin embargo, de que ellos también tienen razones para *rechazar cualquier* contradicción, razones de tanto peso como las que esgrime el dialetheista para *aceptar alguna*: Para un lógico consistente (LC para abreviar), (0') una contradicción es siempre rechazable porque, apelando a principios y leyes de inferencia intuitivamente válidos, es posible demostrar que *de una contradicción se sigue cualquier cosa* (ECQ). La solución que ofrece LC es, pues, satisfactoria porque evita el trivialismo. Aquellos que aceptan la verdad de ciertas contradicciones se ven forzados a rechazar ECQ, ahora bien, esta posibilidad es insatisfactoria para LC por varios motivos: (1') Por restringir la validez general de principios básicos del razonamiento: SD, MP o RA (i.e., *silogismo disyuntivo*, *modus ponens* y *reductio*) que permiten demostrar por separado (en conjunción con leyes válidas también para el dialetheista) la validez de ECQ. (2') Por ofrecer razones *ad hoc* (no independientes de nuestra voluntad de rechazar ECQ) a la hora de restringir SD, MP, RA, etcétera. (3') Por introducir restricciones expresivas (por ejemplo, la expresión de ciertas funciones) y por desdibujar el significado de ciertas conectivas básicas como: '$\neg$' y '$\rightarrow$', al bloquear algunas de las reglas (RA, MP o CTR) que nos ayudan a definir su uso y a describir su comportamiento. (4') Por no ofrecer un criterio que nos permita decidir cuándo una contradicción es una dialetheia y cuándo es sólo falsa, lo que dificulta determinar el significado de "ser coherente" (y también *serlo*) puesto que carecemos de criterios claros para decidir en cada caso cuándo es racional rechazar una contradicción y cuándo no. (5') Porque esta falta de coherencia genera en última instancia "paradojas reforzadas" en el discurso dialetheista que la teoría de Priest no puede resolver por falta de recursos expresivos. (6') Porque todas las contradicciones de inclusión comparten la misma estructura, sin embargo, no todas son dialetheias, así pues, la estructura de las contradicciones de inclusión no determina una postura uniforme ante ellas.

Hemos visto que Priest contempla ya algunas de estas acusaciones (aunque está por ver en qué medida sus respuestas zanjan la cuestión). A vista de pájaro al menos, parece claro que el dialetheismo se enfrenta al mismo tipo de problemas que el resto de soluciones a las paradojas de inclusión (véase 4.2): (1') y (3') cuestionan la capacidad de la teoría de Priest para *conservar* nuestras intuiciones sobre la validez *general* de ciertas leyes de inferencia y sobre el significado de ciertas expresiones básicas; (2') señala el carácter *ad hoc* de muchos aspectos de la teoría; (6') pone en duda el principio de solución *uniforme*; y (4') y (5') acusan al dialetheista de *incoherente*: no hemos erradicado las paradojas, hemos tropezado con otras diferentes. Esto sería particularmente grave porque implicaría, en contra de lo prometido, que el dialetheismo no es una teoría *general* capaz de solucionar *todas* las paradojas sino sólo aquellas que son objeto de su estudio, i.e., las que LP permite expresar. Las paradojas que el dialetheista no puede explicar coherentemente sólo podrían situarse en una especie de metalenguaje o metateoría distinta de LP, lo que no haría sino revivir el fantasma de "la jerarquía" que Priest creía haber exorcizado.

Pese a todo, sólo tenemos de momento un enfrentamiento dialéctico en el que el segundo contrincante (LC), a diferencia del primero (el dialetheista), aporta acusaciones pero no argumentos. El objetivo de la presente sección es nivelar la contienda con argumentos a favor de (0')-(6'). A lo largo de nuestra exposición plantearemos repetidas veces una misma cuestión: "¿Cuándo es una contradicción rechazable y qué contradicciones son rechazables?" Quizá el problema fundamental del dialetheismo a la hora de convencernos se resume en su incapacidad para ofrecer una respuesta satisfactoria a esta pregunta. Por eso en nuestra exposición del problema incidiremos en las causas de ese fracaso y también en sus consecuencias.

5.3.1 Ausencia de criterios sintácticos para detectar dialetheias

Antes de nada, es preciso recalcar que la incapacidad dialetheista para dar una respuesta a la pregunta anterior no es un hecho controvertido:

> "I am frequently asked for a criterion as to when contradictions are acceptable and when not. It would be nice if there were a substantial answer to this question –or even if one could give a partial answer, in the form of some algorithm to demonstrate that some area of discourse is contradiction free. But I doubt that this is possible. Nor is this a matter for surprise. Few would now seriously suppose that one can give an algorithm –or any other informative criterion– to determine when it is rational to accept something. There is no reason why the fact that something has a certain syntactic form –be it $p \wedge \neg p$ or anything else– should change this. One can determine the acceptability of any given contradiction, as of anything else, only on its individual merits."[68]

---

[68] Priest 1998b, p. 423. La misma idea se encuentra también en Priest 1979, IV 9, p. 235; e *IC*, pp. 132-3: "paradoxical sentences do not bear their mark on their sleeve and there seems [to be] no reason to suppose that the class of paradoxical sentences is decidable (i.e., we have no effective way of telling, in general, when a sentence is paradoxical)." (Priest 1979, IV 9, p. 235). De Nuevo: "Given that a theory or hypothesis delivers a contradiction which there are good grounds for rejecting, this provides *prima facie* grounds for rejecting the theory. If there are countervailing reasons one must

Así pues, el dialetheista no puede determinar de antemano cuándo una contradicción es rechazable y cuándo no, se trata de ver y ponderar en cada caso los argumentos (de haberlos) que la apoyan. En este texto, Priest nos indica, además, cuál es, en última instancia, la razón por la que un dialetheista no podría responder a la pregunta formulada: El dialetheista no acepta (y, por tanto, no dispone de) ningún criterio *sintáctico* que identifique oraciones (y, en particular, contradicciones) que debamos *rechazar*. Como todo algoritmo se construye sobre la base de rasgos sintácticos o estructurales compartidos y recurrentes, la ausencia de éstos en el caso de las oraciones rechazables hace que sea imposible determinarlas algorítmicamente. De hecho, dada su complejidad, esta tarea sería inviable aun disponiendo de algunos indicadores sintácticos.

Pese a todo, es incuestionable que la existencia de dichos indicadores haría mucho más sencilla la tarea de decidir cuándo una oración es rechazable. El lógico consistente (LC) tiene en este sentido una ventaja importante, puesto que sí posee *un* criterio sintáctico para determinar de antemano el rechazo de *algunas* oraciones. LC rechaza cualquier oración con *forma* de contradicción: $p \wedge \neg p$, la razón de ello es, como ya sabemos, su creencia en que "de una contradicción se sigue cualquier cosa" (ECQ) y su rechazo del trivialismo. En la medida en que

investigate further. In the end one may decide to accept the inconsistency (though more likely, one will not). [...] If this is all disconcertingly non-algorithmic, this is just an unfortunate fact of life. It is presumably the desire to obtain something more algorithmic el dialetheista acepta algunas contradicciones, ha de renunciar al *único* criterio sintáctico del que disponía LC y, en la medida en que tampoco acepta el trivialismo, ha de renunciar a ECQ, la razón ofrecida por LC para proscribir las contradicciones. Sin embargo, esta decisión tiene consecuencias graves.

5.3.2 Restricciones inferenciales

Vimos (5.2.2.1, 5.2.3.1) que rechazar ECQ nos empuja a rechazar la validez general de todos aquellos principios que (junto a principios válidos para el dialetheista) nos permiten inferir ECQ, entre ellos las leyes de *silogismo disyuntivo* y *modus ponens*. Estas restricciones empobrecen intolerablemente nuestras prácticas inferenciales. Priest es consciente de ello y ofrece una solución doble (véase 5.2.3.2 y 5.2.3.3) consistente en (a) introducir nuevas conectivas que reflejen nuestra noción intuitiva de implicación (y que cumplan MP, etc.); y (b) aceptar la validez de todos los principios clásicos en contextos *consistentes*, i.e., aquellos que no involucran dialetheias. Priest reconoce que la primera estrategia se ha mostrado insuficiente a la hora de preservar todas las inferencias clásicas que el dialetheista desea preservar.[69] Por ello Priest insiste en la segunda y declara "cuasi-válidas" a las reglas cuya validez sólo se suspende en presencia de dialetheias. Esta estrategia persigue que, en la práctica, las inferencias realizadas por un lógico clásico y un

that is behind the demand that all contradictions should be rejected. Such a demand cannot be met. Nor there is any reason why it should be." *IC*, p. 132.

[69] Priest 1991, pp. 321-2. Véase nota 57.

dialetheista sean indiscernibles *en contextos consistentes.*[70] Sin embargo, el éxito de esta estrategia depende crucialmente de que podamos identificar los contextos consistentes y distinguirlos de los inconsistentes o, en otras palabras, de que podamos decidir si una contradicción es una dialetheia o una simple falsedad rechazable. En resumen, sólo podremos asegurar que nuestras prácticas son indistinguibles de las del lógico clásico (en contextos consistentes) cuando seamos capaces de responder satisfactoriamente a la pregunta que formulamos en 5.3. Priest (1991) nos muestra que la idea de una lógica mínimamente inconsistente, LPm, tiene sentido, pero, como señalábamos en la nota 59, *cualquier* contradicción nos remite en LPm a modelos inconsistentes y suspende, por tanto, el uso de reglas cuasi-válidas. Dado que la probabilidad de que una contradicción sea una dialetheia es ínfima, recurrir a modelos inconsistentes en presencia de *cualquier* contradicción restringiría seriamente las pretensiones de Priest de equiparar las inferencias de LPm con las de la lógica clásica en contextos consistentes.

[70] "Suppose [...] that a contradiction never turns up. Then both the classical and the dialetheic logician can accept their (quasi-)proofs at face value. Both will accept exactly the same consequences of the theory. Thus, in consistent situations classical consequence can be understood in terms of quasi-validity: the dialetheic logician is able to reason in exactly the same way that the classical logician does." *IC*, p. 147. Más adelante añade: "dialetheic logic gives the full power of classical logic except where classical logic is demonstrably useless, and then more." *IC*, p. 148. Véase, en general, *IC*, 8.5 o Priest 1991. En Priest 1994b (pp. 340-1) se ofrece un argumento similar con respecto a las operaciones aritméticas que efectúa en la práctica un matemático que opera con modelos consistentes de la aritmética y un dialetheista.

5.3.3 La idea de absurdo y el significado de negar: Negación y rechazo (I)

Además de las restricciones inferenciales mencionadas, el abandono de ECQ y, por consiguiente, de un criterio *sintáctico* para identificar oraciones rechazables –aun cuando no toda oración rechazable satisfaga dicho criterio–, dificulta (aunque no imposibilita) dos tareas distintas pero íntimamente relacionadas: (a) Identificar una estrategia general para rechazar creencias, afirmaciones, etc.; y (b) explicar el significado del operador lógico de negación: '¬'. Con respecto a (a), el lógico consistente, LC, puede razonar del siguiente modo: "Aunque una oración, p, no tenga la forma: $q \wedge \neg q$, sería posible rechazar su contenido si suponer (la verdad de) p implicase una contradicción". Del mismo modo, LC puede añadir: "Si rechazamos el contenido de p, entonces negamos p (i.e., aceptamos $\neg p$), así pues, si p implica una contradicción, entonces $\neg p$". Este último hecho es relevante con respecto a (b) porque, si entendemos que fijar el significado de un operador '#' consiste en aportar reglas generales que determinen su uso y, en particular, su introducción (I#) y su eliminación (E#), entonces LC dispone de criterios sintácticos para definir las reglas de la negación, '¬':

$$\dfrac{\begin{matrix}[A] \\ B \wedge \neg B\end{matrix}}{\neg A}\ (I\neg) \qquad\qquad \dfrac{\neg\neg A}{A}\ (E\neg)$$

Según (I¬): "Si suponer A implica una oración del tipo, B ∧ ¬B, entonces podemos inferir ¬A". Según (E¬): "Si una oración tiene la forma ¬¬A, entonces podemos inferir A". El lógico dialetheista no tiene ningún problema con (E¬), ya que acepta la ley de doble negación (véanse 5.2.2.3 y nota 38), pero no ocurre lo mismo con (I¬). Si B es una dialetheia, entonces el hecho de que A implique B $\wedge$ $\neg$B no motiva que podamos negar (o rechazar el contenido de) A.

Tanto la estrategia ofrecida para rechazar enunciados que no son contradicciones como la regla (I¬) de introducción de la negación, presuponen dos cosas: (I) el principio de *reductio ad absurdum* (RA) según el cual "si de una oración, A, *se sigue un absurdo*, entonces debemos rechazarla/negarla"; y (II) la tesis de que "*toda* contradicción es un absurdo". El dialetheista debe rechazar (I) o (II), ya que su conjunción permite demostrar ECQ, como vimos en la nota 27. Rechazar (I) es inviable, de otro modo nuestra comprensión de lo que significa "rechazar" o "negar" algo se vería seriamente afectada. Rechazar o negar *p* significa en buena medida que *p* es inaceptable o que tiene consecuencias inaceptables. Un "absurdo" es algo racionalmente inaceptable, así pues, todo aquello que sea absurdo o que conduzca a un absurdo debe ser rechazado y negado, como afirma RA. Para algunos, un "absurdo" es además algo devastador de lo que se sigue la verdad de todo (el trivialismo), pero, aceptemos o no esta tesis fuerte, el resultado es invariable: un absurdo es siempre inaceptable. Parece obvio, pues, que un dialetheista está obligado a respetar (I).

El rechazo de (II), en cambio, está perfectamente justificado para un dialetheista, después de todo, para él las dialetheias no son absurdas. No es irracional creer en la verdad de ciertas contradicciones basándonos en los argumentos que las sustentan (recordemos también la paradoja del prólogo). Sólo sería irracional creer una contradicción si supiésemos que ésta es falsa sin más o (dada la baja probabilidad de las dialetheias) si no tuviésemos motivos para creer en su verdad. Ahora bien, si una contradicción no es siempre un absurdo, RA no puede formularse como sugiere LC: "Si de A se sigue una contradicción, entonces debemos rechazar/negar A". Si B $\wedge$ $\neg$B es una dialetheia y, por ello, verdadera, el hecho de que A implique B $\wedge$ $\neg$B no justifica ni el rechazo ni la negación de A, ya que no tenemos razones para pensar que A sea falsa o no verdadera.

Cuando el dialetheista quiere ofrecer una estrategia general para rechazar oraciones o una regla que nos permita decidir cuándo introducir el operador de negación, '$\neg$', sólo tiene que apelar a RA sin identificar absurdo con contradicción: "Algo es rechazable/negable si la suposición de su verdad conduce a un absurdo". A su vez, la regla de introducción de la negación, (I$\neg$), se enuncia ahora del siguiente modo:

$$\frac{\begin{array}{c}[A]\\ \bot\end{array}}{\neg A}\ (\mathrm{I}\neg)$$

donde '$\bot$' representa la idea de absurdo o falsedad, pero 'B $\wedge$ $\neg$B $\leftrightarrow$ $\bot$' no se cumple en general (ni tampoco 'B $\wedge$ $\neg$B $\rightarrow$ $\bot$').[71]

[71] Priest (1999, pp. 112-3) no sólo muestra cómo es posible para un dialetheista apropiarse de RA o (I$\neg$), ¡también muestra cómo apropiarse de ECQ! Para ello, eso sí, apela a una noción de negación *distinta* de (aunque *compatible* con) la negación estándar dialetheista '$\neg$'. Esta nueva negación, '–', recibe el nombre de "*arrow falsum*" y se define del siguiente modo: $-p$ sii $p \rightarrow \bot$ (donde '$\rightarrow$' es un condicional (no material) que cumple *modus ponens* y '$\bot$' se interpreta como $\forall x$ V(x), i.e., "todo es verdadero", y, por tanto, se cumple (por PEV) que: $\bot \vdash q$, para todo q). Está claro que: $p, -p \vdash q$

En la medida en que los críticos de Priest defienden que cualquier contradicción es absurda, sus argumentos buscan mostrar que las ideas de rechazo racional y negación pierden sentido sin la equivalencia: $B \wedge \neg B \leftrightarrow \perp$ (para *todo* B). Tennant (1998, pp. 31-4), por ejemplo, sostiene que la idea de negación deriva de la constatación de que ciertas proposiciones mantienen entre sí una relación de *contrariedad necesaria* que las hace *incompatibles*. Ni siquiera las proposiciones atómicas son lógicamente independientes unas de otras en su evaluación. La verdad de (A) "Estoy en Barcelona" es incompatible con la verdad de (B) "Estoy en Valencia" referidos al mismo individuo, lugar y momento. Creer la verdad de ambas en esas condiciones es absurdo. Para Tennant, negar A, por ejemplo, *significa* que suponer su verdad en un argumento entre cuyas premisas, X = {B, ...}, se encuentra algo que (como B) es necesariamente contrario a A conduce a un absurdo. ¬A significa que A es incompatible con la verdad de (alguna de) las premisas contenidas en X:

$$\frac{\begin{array}{c}X, [A]\\ \vdots \\ \perp\end{array}}{\neg A}\ (\mathrm{I}\neg)$$

Tennant parece sugerir que rechazar PNC violaría el *significado* de la negación por ser la contradicción el caso extremo de contrariedad necesaria. ¬A *significaría* entre otras cosas: Aquello que es necesariamente incompatible con (o contrario a) la verdad de A.[72]

(i.e., $p, p \rightarrow \perp \vdash q$), ahora bien, si p es una dialetheia, $v(p) = \{0, 1\}$, se cumple $p \wedge \neg p$, pero *no se cumple* $p \wedge -p$, por tanto: $p \rightarrow \perp = -p \neq \neg p$.

[72] "Denying the law of non contradiction [...] seems to me a desperate meaning violating move. It is desperate because there are surely better responses to the paradoxes [...]. It is

Priest (*BLT*, pp. 273-4) responde a Tennant que un lógico dialetheista no tiene por qué aceptar su teoría de la negación y añade otras críticas: 1) En la teoría de la negación de Tennant (1999), ECQ tampoco es válido y, por tanto, no supone una crítica a la lógica paraconsistente de Priest; 2) Tennant no explica lo que significa negar una proposición en ausencia de premisas incompatibles, i.e., cuando de la mera suposición de A se sigue $\perp$ y, por tanto, $\neg$A; 3) la interpretación que da Tenannt a '$\perp$' ("absurdo metafísico") es demasiado fuerte, negar que estoy sentado no es afirmar que "estar sentado" sea un absurdo metafísico. Las réplicas de Priest son claramente insatisfactorias. Si en un cálculo deductivo inferimos $\neg$A en ausencia de otras premisas, es porque A es contrario a algún *axioma* o *principio* que no hace falta especificar entre las premisas y, por tanto, rechazable sobre la base de su incompatibilidad con la verdad de otras proposiciones, aunque éstas no figuren como premisas explícitas en el argumento. Por otra parte, a partir de la interpretación que hace Tennant de $\perp$ no se sigue que sea metafísicamente absurdo "estar sentado" cuando uno está de pie, sino sólo que lo sea "estar sentado y de pie al mismo tiempo". Por último, quizá la lógica empleada por Tennant sea paraconsistente, pero esto es irrelevante, un lógico consistente podría afirmar que parte de lo que '$\neg$A' *significa* es "ser incompatible con (en el

---

meaning violating because of what negation means. One proves a negation $\neg$A by showing that the assumption of (the truth of) A is contrary to (the assumed truth of) whatever other premises X one wishes to hold on to [...]. Thus, I would argue, the notion of negation actually rests on the more fundamental notion of *necessary contrariety*, or *impossibility of join truth* of at least two different propositions. For the notion of negation to have entered the language it would have sufficed, on this view, for the language to have *contrary* atomic propositions." (Tennant 1998, pp. 33-4) sentido de excluir) la verdad de A". Pese a todo, la primera observación de Priest es correcta, un dialetheista *no tiene por qué* aceptar la teoría de la negación de Tennant. El dialetheista podría seguir en sus trece e insistir en que, A y ¬A no son siempre incompatibles (aunque lo sean habitualmente), así pues, no es verdad que '¬A' *signifique en general* "ser incompatible con A".[73]

Autores como Weir (1999, pp. 124-5) apuntan por su parte que el dialetheismo imposibilita nuestra comprensión de lo que supone el abandono racional de una teoría. Según Priest, una teoría se puede rechazar racionalmente cuando entraña algo muy poco probable (por ejemplo, una contradicción), a menos que tengamos pruebas de que "lo poco probable" ha sucedido esta vez. Según Weir, la noción de probabilidad se puede entender aquí de dos modos: o bien como la relación *objetiva* que mantienen entre sí ciertas proposiciones, de tal modo que la probabilidad de que q sea verdad cuando lo son $p_0, ..., p_n$ sea, *de hecho*, x, para $0 \leq x \leq 1$; o bien como algún tipo de relación entre proposiciones establecida *subjetivamente*. En el primer caso, Priest

[73] Zalta (véase *BLT*, pp. 271-2) ofrece un argumento a favor de PNC centrado en la ejemplificación de propiedades. Rechazar el principio de no contradicción hace ininteligible nuestra idea intuitiva de "ejemplificación de una propiedad". Sólo existen *dos* posibilidades incompatibles entre sí: a logra satisfacer P sii P(a); a fracasa en su intento por satisfacer P sii ¬P(a). Aceptar P(a) y ¬P(a) hace que nuestra comprensión de la idea de ejemplificación pierda sentido. Priest (*BLT*, p. 272) responde que el problema del dialetheismo afecta fundamentalmente a la *negación* no a la ejemplificación de propiedades y que ambos problemas sólo pueden vincularse si nuestra teoría de la negación acepta: $\neg p$ es verdadero sii p no es (i.e., p fracasa en su intento por satisfacer la propiedad de ser) verdadero. Ahora bien, el dialetheista no acepta tal equivalencia (ya que en su teoría de la negación p y $\neg p$ pueden ambos ser verdaderos), por tanto, suponerla sin ofrecer argumentos a su favor es incurrir en una petición de principio contra Priest.

abraza un proyecto que ha fracasado históricamente (el de establecer relaciones objetivas de probabilidad entre proposiciones). En el segundo, nada nos impide (dado que el dialetheista *permite* creer algunas contradicciones) considerar que la verdad de una contradicción implicada por una teoría que nos parece *subjetivamente* aceptable tiene una probabilidad de 1. En otras palabras, si una teoría inconsistente nos gusta mucho, creeremos que las contradicciones que genera tienen probabilidad 1, aunque la teoría sea incorrecta y sus contradicciones falsas.

Priest (*BLT*, p. 275) rechaza este argumento insistiendo en lo que ya dijimos con respecto al rechazo de una contradicción. Las probabilidades que una proposición, *p*, tiene de ser verdadera son *revisables* y el dialetheista siempre tiene la opción de aceptar o rechazar cualquier teoría o proposición, *p*, sobre la base de una investigación metódica que sopese las pruebas disponibles a favor y en contra de *p*. Si aparecen nuevas proposiciones cuya verdad es altamente probable (improbable) y de las que podemos inferir *p*, entonces revisamos al alza (a la baja) las probabilidades que tiene *p* de ser cierta. Obviamente, este tipo de investigaciones es siempre revisable. Quejarnos de que ninguna investigación empírica de este tipo podría cuestionar un principio como PNC, supondría dar por sentado PNC e incurrir en una petición de principio contra el dialetheista.

El dialetheista tiene, pues, un argumento estándar para hacer frente a quienes le acusan de vaciar de contenido o significado las nociones de negación y rechazo: Basándose en RA, un dialetheista puede negar o rechazar todo aquello que entrañe un absurdo. En este sentido, recoge las intuiciones que todos tenemos con respecto a "rechazar" y "negar". El dialetheista *sólo* se opone a la idea de que toda contradicción sea absurda. La fuerza de su argumento (y esto es vital) descansa por completo en la existencia de paradojas y en la tesis de que es imposible resolverlas de forma consistente. Para sacar de su "error" al dialetheista, un lógico consistente debería mostrar de forma satisfactoria que todas las contradicciones se pueden negar o rechazar consistentemente, lo que implicaría resolver las paradojas de inclusión. Así pues, todo aquel que insista en la validez universal de PNC debe abordar con éxito esta tarea, de lo contrario incurre en una petición de principio.

5.3.4 Falso y no verdadero: Negación y rechazo (II)

En el anterior apartado, el lógico consistente, LC, argumentaba que p y $\neg p$ son incompatibles, se excluyen mutuamente. Para ello apelaba al significado de "negar" y "rechazar" y al modo en que ambas nociones se relacionan con la idea de "absurdo", defendiendo la equivalencia: $B \wedge \neg B \leftrightarrow \perp$ (o, al menos, $B \wedge \neg B \rightarrow \perp$), para todo B. Al igual que en 5.3.3 examinaremos ahora argumentos que defiendan la incompatibilidad entre p y $\neg p$, pero lo haremos explotando las conexiones que existen entre las nociones de negación y rechazo y las de falsedad y no-verdad.

Hemos estado hablando de "negación" [*negation*] y "rechazo" [*denial*], conviene recordar ahora que ambas nociones son diferentes (véase el final de 5.2.2.3). Uno puede rechazar un enunciado, p, por diversos motivos relacionados con las intenciones del hablante y con los presupuestos que condicionan la comunicación (*p* podría ser inapropiado, o irrelevante, o engañoso, etc.). Pero uno también puede rechazar *p* por motivos estrictamente semánticos, a saber, porque el contenido de *p* –la información que comunica– no refleja ningún hecho real y, por tanto, *p* no es verdadera sino falsa. En el primer tipo de situación rechazar un enunciado *no implica negar su contenido*, en el segundo, en cambio, sí.[74] El rechazo que nos ocupa (tanto ahora como en el apartado anterior) es el que se dirige al *contenido de un enunciado* o a *su verdad*. Desde el punto de vista tradicional (clásico y consistente), esta forma de rechazo se encuentra indisolublemente ligada a la operación semántica de negación: *Rechazar* (*el contenido de*) *p es equivalente a aceptar* $\neg$ *p* (i.e., *negar p*), donde equivalente significa aquí que, *quien haga una de estas cosas, "se compromete racionalmente" a hacer la otra*.[75] LC objeta al dialetheista que no puede dar cuenta de esta equivalencia, ya que cuando *p* es una dialetheia cabe aceptar $\neg$ *p sin rechazar p*, lo que resulta imposible.

Como respuesta a esta acusación, Priest trata de cuestionar la validez de la equivalencia supuesta por LC, para ello traza dos contrastes entre los pares "*p* y negación de *p*", por un lado, y "aceptación de *p* y rechazo de *p*", por otro:

(I) Las categorías de aceptación y rechazo *no son exhaustivas*: No aceptar *p*, no implica rechazar *p*; y no rechazar *p* no implica aceptar *p*. Si uno es

[74] La oración 'No me llamo Anacleto' podría ser usada por alguien, no para negar que 'Anacleto' sea su nombre, sino para rechazar ser llamado así (por preferir un diminutivo, otro nombre, o por otro motivo). Obsérvese, así mismo, que las acciones de "negar" y "rechazar sin negar" mediante una oración *p* no dependen de la presencia en *p* de una palabra concreta como 'no' sino de factores contextuales.

[75] Véase Smiley 1993, pp. 20-2.

agnóstico con respecto al contenido de p, entonces ni acepta ni rechaza p. Sin embargo, con p y su negación no ocurre lo mismo. Si p no se da, se da su negación, $\neg p$, y si la negación de p no se da (i.e., si se da $\neg\neg p$), entonces se da p. Esto está relacionado con el hecho de que las categorías "verdad" (V) y "falsedad" (F) *son exhaustivas* ($\neg V(\underline{p}) \rightarrow F(\underline{p})$ y $\neg F(\underline{p}) \rightarrow V(\underline{p})$, para todo p)[76] y con la decisión de definir 'p es falso' en términos de '$\neg p$ es verdadero' ($F(\underline{p})$ sii $V(\underline{\neg p})$ sii $\neg p$). Si p no se da (i.e., $\underline{p}$ no es verdadero), entonces $\neg p$:

$\neg V(\underline{p}) \Rightarrow_{\text{exhaust.V/F}} F(\underline{p}) \Rightarrow_{\text{def.}} V(\underline{\neg p}) \Rightarrow_{\text{PEV}} \neg p$

Si $\neg p$ no se da (i.e., $\underline{\neg p}$ no es verdadero), entonces p:

$\neg V(\underline{\neg p}) \Rightarrow_{\text{exhaust.V/F}} F(\underline{\neg p}) \Rightarrow_{\text{def.}} V(\underline{\neg\neg p}) \Rightarrow_{\text{PEV}} \neg\neg p \Rightarrow_{\text{PDN}} p$

(II) El segundo contraste señala que las categorías de aceptación y rechazo son mutuamente *excluyentes*, es imposible aceptar y rechazar p al mismo tiempo, el rechazo y la aceptación de p son incompatibles desde la perspectiva *psicológica* de un individuo.[77] Sin embargo p y su negación, $\neg p$, *no se excluyen* necesariamente, las categorías de verdad y falsedad no son mutuamente excluyentes. Es posible que p sea verdadero y falso y, por tanto, que se den a la vez p y $\neg p$. Este contraste muestra que negar y

---

[76] Debido a que la ley de silogismo disyuntivo ($\neg A, A \vee B \models B$) no tiene validez general para el dialetheista, la exhaustividad de p y $\neg p$ (o de $V(\underline{p})$ y $F(\underline{p})$) no puede expresarse mediante la disyunción: $p \vee \neg p$ (o $V(\underline{p}) \vee F(\underline{p})$), sino mediante las implicaciones $\neg p \rightarrow \neg p$ y $\neg\neg p \rightarrow p$ (o $\neg V(\underline{p}) \rightarrow F(\underline{p})$ y $\neg F(\underline{p}) \rightarrow V(\underline{p})$), '$\rightarrow$' se debe entender, además, como un condicional *no* material que respeta la ley de *modus ponens*.

[77] "Acceptance and rejection are not exhaustive, but they are exclusive. They are not exhaustive since being agnostic is a third possibility. In particular, therefore, rejecting something is not the same as not believing it: it is something much stronger. On the other hand acceptance and rejection do appear to be incompatible. One can certainly believe something and believe its negation. [...] But it seems difficult to argue that one might both believe something and *refuse* to believe it." *IC*, pp. 122-3.

rechazar (incluso cuando se trata de rechazar *el contenido* de un enunciado) son categorías distintas y no equivalentes: aceptar *p* no excluye negar *p* pero sí rechazar *p*.

Priest tiene razón al afirmar que "negación" y "rechazo" son cosas distintas, pero cabe señalar, no obstante, que todo lo que se dice en (I) es compatible con la equivalencia defendida por LC. "Rechazar" y "negar" pueden ser cosas diferentes y, pese a todo, *equivalentes* en el sentido especificado.[78] Así pues, (I) no cuestiona la tesis de LC, (II), en cambio, sí. (II) afirma que la negación de *p no nos compromete* con su rechazo, una vez más, el argumento se basa en la posibilidad de aceptar p y $\neg p$, i.e., en la tesis básica del dialetheista según la cual *p* podría ser verdadera y falsa. El principal apoyo de esta tesis es, como ya sabemos, la inexistencia de una solución consistente satisfactoria para las paradojas (el molesto pero imbatible argumento de siempre). Así pues, el desencuentro fundamental afecta a la implicación: "negar *p* implica rechazar (el contenido de) *p*", el lógico clásico acepta esta implicación, el dialetheista no. Ambos bandos están de acuerdo, no obstante, en que "rechazar (el contenido de) *p* implica negar *p*".

Pero las cosas no son tan sencillas para el dialetheista. El lógico clásico aún puede aportar nuevos argumentos a favor de la equivalencia entre negar y rechazar: Negar *p* equivale a afirmar que $\neg p$ es verdadero y también a afirmar que *p* es falso [$\neg p$ sii $V(\underline{\neg p})$ sii $F(\underline{p})$]. Ahora bien, si *p* es falso, entonces no es verdadero [$F(\underline{p})$ sii $\neg V(\underline{p})$] y es, por tanto, rechazable. Este argumento es válido para LC por dos motivos: (A)

Porque falsedad y no-verdad *son lo mismo* [F(p) $=_{def}$ ¬V(p)]; y (B) porque, desde un punto de vista epistemológico, parece racional aceptar lo que es verdad y *rechazar lo que no es verdad* (o lo que es *falso*). Priest afirma, de hecho, que el objetivo de la aserción o de la aceptación racional es la verdad (véase 5.2.2.2), así pues, parece apropiado pensar en la existencia de un principio dual según el cual el objetivo del rechazo racional es la no-verdad o la falsedad. *Si p no es verdadero* (y es, por tanto, falso), *entonces debe ser rechazado.* LC respeta esta intuición básica, el lógico dialetheista, en cambio, no. Aceptar una dialetheia es aceptar una falsedad tanto como aceptar una verdad, sin embargo, *sólo* las verdades son aceptables, las falsedades son rechazables. Sería inútil replicar que las dialetheias son aceptables y rechazables al mismo tiempo porque, como reconoce el propio Priest, el rechazo y la aceptación de *p* son psicológicamente incompatibles.[79]

Priest niega los dos puntos de apoyo de este argumento. Contra (A) afirma que falsedad y no-verdad *no son* equivalentes. Contra (B) viene a decir que el objetivo del rechazo racional no es lo no-verdadero sino "lo no-verdadero *que no es verdadero*". Ambas respuestas resultan *ad hoc* e insatisfactorias. Con respecto a (A), Priest acepta que ¬V(p) implica F(p) (por la "exhaustividad" de V y F), pero rechaza que F(p) implique ¬V(p). Así pues, falsedad y no-verdad son categorías *distintas* y *no equivalentes*. Un enunciado puede ser falso y no ser no-verdadero. Pero, ¿es realmente posible distinguir lo no-verdadero de lo falso? Priest

---

[78] Véase Smiley 1993, p. 20.
[79] Véase por ejemplo Smiley 1993, p. 22.

se ve obligado a introducir esta distinción fundamentalmente por razones (no ya de consistencia sino) de *coherencia*. Cuando el dialetheista afirma que una oración, p, puede ser verdadera y falsa, *no niega* la verdad de p, al contrario, *la afirma*. Si "falsedad" y "no-verdad" fuesen equivalentes, entonces decir que p es verdadera y falsa sería como afirmar y *negar* al mismo tiempo que p es verdadera. Ahora bien, en la medida de lo posible, el dialetheismo *no desea negar* la verdad de una dialetheia. Priest trata de evitar este tipo de contradicciones "(meta)teóricas", pero resulta imposible impedir que se produzcan algunas en el corazón mismo de su teoría. A partir de la paradoja del mentiroso, i.e., $\underline{\mu}$ = '$\neg V(\underline{\mu})$' (unida a PEV y PTE, principios ambos aceptados por el dialetheista), podemos afirmar, no sólo que $\underline{\mu}$ es verdadera y falsa, sino también que $\underline{\mu}$ *no es verdadera* (pese a ser una dialetheia). Es más, Priest admite que podemos demostrar: (1) $\exists x(V(x) \wedge \neg V(x))$ y (2) $\neg\exists x(V(x) \wedge \neg V(x))$ (véase el final de 5.2.2.2). Si falsedad y no-verdad fuesen equivalentes, entonces podríamos demostrar también: (1') $\exists x(V(x) \wedge F(x))$ y (2') $\neg\exists x(V(x) \wedge F(x))$. Mientras que Priest no tiene ningún inconveniente, siendo dialetheista, en aceptar (1) y (1'), aceptar (2) y (2') resulta algo incómodo porque lo que ambos enunciados dicen de forma intuitiva es que verdad y no-verdad, en el caso de (2), y verdad y falsedad, en el de (2'), son categorías mutuamente excluyentes. Esta afirmación *niega* y, por tanto, hace *falsa* la tesis central del dialetheismo: "Hay enunciados verdaderos y falsos". Priest insiste a menudo (véase 5.3.5 y las notas 36, 80 y 92) en que una inconsistencia "metateórica" no invalidaría la teoría dialetheista, en primer lugar porque el dialetheismo trata de justificar la verdad de ciertas contradicciones y, en segundo, por su rechazo de la distinción "teoría-metateoría". Ambas son una misma cosa para el dialetheista, por tanto, asumir que la teoría contiene dialetheias es asumir que *todo* el entramado teórico las contiene.[80] De este modo, Priest acaba por admitir (1'), pero se resiste a aceptar (2') alegando que no hay motivo para multiplicar innecesariamente las dialetheias y que no hay razones para aceptar (2'), una vez hemos distinguido entre falsedad y no-verdad. La dicotomía "verdad/no-verdad" genera así más inconsistencias que la dicotomía "verdad/falsedad": (1) y (2) establecen que las categorías "verdad/no-verdad" *son y no son* mutuamente excluyentes, aunque, según Priest (*IC*, p. 91), esto sólo afecta a oraciones relacionadas con las paradojas.[81]

¿No es éste acaso un movimiento *ad hoc* encaminado a salvar la coherencia del dialetheismo y a evitar que su tesis central sea ella misma una dialetheia y, por tanto, *falsa* (además de verdadera)? No existen

[80] "I am affirming both [(1)] and [(2)]. This is, of course, a contradiction. [...] If I were attempting to produce a consistent theory of the consistent, this would be fatal. However, the aim of the enterprise is not to eliminate contradictions but to accommodate them." (*IC*, p. 91)

[81] Es interesante señalar que Priest cambia de opinión con respecto a (1') y (2') entre 1979 y 1987: "though it is easy enough to argue for $\exists x(\mathrm{T}(x) \wedge \mathrm{F}(x))$ (the analogue of [(1)]) on the basis of the liar paradox, it seems impossible to argue for $\neg\exists x(\mathrm{T}(x) \wedge \mathrm{F}(x))$ (the analogue of [(2)]), at least without the exclusion principle." (*IC*, p. 91). Compárese con: "consider the metalinguistic statement [(1')] Some sentences are true and false (i.e. '$\exists s$(s is true and s is false)' where the quantifier ranges over all true or false sentences –which, of course, includes paradoxical ones). Then using the above tables [K3F, nota 45] and the truth conditions for quantifiers [...], [(1')] can be seen to be true, in fact paradoxical. Thus its negation [(2') No sentence is true and false] is true too. Both my claim that there are paradoxical sentences and Aristotle's claim [...] that there are none are true!" (Priest 1979, pp. 238-9). La diferencia entre ambos textos es que en 1979 Priest no separa falsedad de no-verdad y utiliza tres valores de verdad: t, f, p, en vez de dos: 0, 1, combinados con tres posibles evaluaciones semánticas: $\{0\}$, $\{1\}$, $\{0,1\}$.

motivos ajenos a la justificación del dialetheismo que nos lleven a distinguir falsedad y no-verdad. Esto resulta insatisfactorio, entre otras cosas, porque todas las prácticas semánticas o epistémicas que aquí nos preocupan con respecto al contenido de una oración o proposición *p* (negación, aceptación, rechazo o agnosticismo) se pueden describir centrándonos en la presencia o ausencia de la propiedad 'ser verdadero'. Suponer que falsedad y no-verdad son nociones equivalentes es la hipótesis más plausible y la que mejor recoge nuestras intuiciones. El dialetheismo debería justificar de forma independiente su distinción entre ambas nociones.[82] Si Priest está dispuesto (como él mismo afirma) a aceptar que ciertas tesis dialetheistas podrían ser dialetheias, entonces no existe *en principio* ningún inconveniente para identificar falsedad y no-verdad y asumir que la tesis central del dialetheismo: "Hay enunciados que son verdaderos y falsos" es una dialetheia y, por tanto, falsa (además de verdadera). Si Priest evita esta ecuación es porque, en el fondo, no resulta agradable asumir que los enunciados centrales del dialetheismo son dialetheias. Para evitarlo hace que las nociones utilizadas para articular su teoría (verdad y falsedad) se comporten consistentemente *en*

[82] La crítica es de Sainsbury (1995, en la 1ª edición de 1987: pp. 143-4): "There is no room for an intelligible conception of falsehood that differs from nontruth. All our semantics, and a proper account of the conditions for rational belief, can be effected in terms of a single property, truth, which may be present or absent, but not both present and absent. To recognize its presence is to accept. To recognize its absence is to reject. To fail to recognize its presence and fail to recognize its absence is to be agnostic. Negation turns truth into nontruth, nontruth into truth. Conjunctions with nontrue conjuncts are nontrue. Nothing can be both true and nontrue. There is no room for dialetheism, since there is no room for a distinction between nontruth and falsity. [...] The dialethist now has a debt to pay: He must justify the distinction between nontruth and falsehood. This means explaining why the distinction is required for purposes other than the defense of dialetheism."

la exposición de la misma y relega las contradicciones al binomio verdad/no-verdad mediante una distinción *ad hoc* entre falsedad y no-verdad.

Priest también se opone a la idea expresada en (B) según la cual el objetivo del rechazo racional es lo no verdadero (o lo falso). Según él, no es racional rechazar en general las cosas que parecen no-verdaderas (o falsas), lo racional, debido a las paradojas, es rechazar aquellas cosas que *parecen no-verdaderas* (o que *parecen falsas*) y que, además, *no parecen verdaderas*. Es decir, sólo debemos rechazar una creencia, *p*, cuando *tenemos indicios de que p no es verdadera y carecemos de indicios de que p es verdadera*, de lo contrario *p* podría ser verdadera y no ser verdadera.[83] Tal vez esta última afirmación "suene mejor" si la expresamos diciendo que *p* podría ser verdadera y no-verdadera, sin embargo distinguir formalmente entre "ser no-verdadero" y "no ser verdadero" resultaría *ad hoc* y complicaría las cosas intolerablemente, tendríamos entonces tres formas negativas no equivalentes: "ser no-verdadero", "ser falso" y "no ser verdadero" y tendríamos que explicar, además, cómo se relaciona la negación, '$\neg$', con todas ellas, especialmente con la primera. Así pues, no es incorrecto decir que las expresiones '*p* es no-verdadera' y '*p* no es verdadera' son equivalentes, al

[83] "[T]o say that untruth is the *telos* of rejection is to say that if a sentence appears to be untrue (in the light of all the evidence, etc.), one ought to reject it. Now, why should one suppose this? [...] It may seem plausible when one think of run-of-the-mill examples. But what of those singular situations, such as the one given by $\underline{\lambda}$ $[\lambda \leftrightarrow \neg V(\underline{\lambda})]$, where one seems to have evidence that a sentence is *both* true and untrue. Is it right to reject a claim of this kind? Not obviously. More plausibly, one should reject a statement if it appears to be untrue and does not appear to be true. That is the *telos* of rejection." (Priest 1993, p. 40)

igual que las afirmaciones: "Sólo debemos rechazar las oraciones no-verdaderas cuando no son verdaderas" y "Sólo debemos rechazar las oraciones que no son verdaderas cuando no son verdaderas". Ahora bien, si la segunda afirmación no tiene mucho sentido (por obvia), es poco probable que la primera tenga más. El uso del adjetivo negativo 'no-verdadero' es lo que genera la ilusión de inteligibilidad al mostrar las expresiones "ser no-verdadero" y "no ser verdadero" como distintas cuando, en lo sustancial, no lo son.[84]

Priest podría protestar y puntualizar que el objetivo del rechazo no es "lo no-verdadero que no es verdadero", sino "aquello de cuya no-verdad o falsedad *tenemos indicios* y de cuya verdad *carecemos de indicios* (i.e., lo que *nos parece* no-verdadero o falso y, además, *no nos parece* verdadero)". De nuevo, no está claro que "parecer no-verdadero" y "no parecer verdadero" sean cosas distintas, aunque sí es cierto que una cosa nos puede parecer a la vez verdadera y no verdadera por tener indicios a favor de ambas apariencias (como ocurre en las paradojas). Sin embargo, la puntualización de Priest *es incorrecta*. Ni el objetivo del rechazo es lo que *nos parece* no-verdadero (o falso, omitiré "falso" en lo sucesivo) ni el objetivo de la aceptación es lo que *nos parece* verdadero. El objetivo del rechazo es *lo no verdadero* y el de la aceptación *lo verdadero*. Es cierto que sólo podemos aceptar lo que *nos parece* verdadero y rechazar lo que *no nos parece* verdadero (y que las

[84] Priest podría optar por una formulación menos chocante: "Debemos rechazar las oraciones que son falsas y que no son verdaderas", pero esto sería insuficiente. Como vimos antes, $\underline{\mu}$ = '$\neg V(\underline{\mu})$' *es falsa y no es verdadera*, sin embargo Priest debe aceptarla porque (además de falsa y no verdadera) $\underline{\mu}$ *es verdadera.*

apariencias engañan), pero esto es así debido a nuestras limitaciones cognitivas, no a que el objetivo de la aceptación o del rechazo sea lo aparentemente verdadero/no-verdadero. No podemos cuestionar la racionalidad de alguien que rechaza lo que no le parece verdadero, ¿qué otra cosa podría hacer? Sin embargo, es correcto criticar la racionalidad de alguien que cree que debemos rechazar algo simplemente porque *no nos parece verdadero* (o porque *nos parece no-verdadero*) y no porque *no sea verdadero*. Si nuestra obligación racional fuese rechazar lo que no nos parece verdadero y no lo no-verdadero sin más, un niño que aprende a hacer operaciones aritméticas y que rechaza '2+2 = 4' porque, de hecho, *no le parece* verdadero, *habría hecho lo correcto*. Pero esto no es así. Rechazar '2+2 = 4' *no es correcto en ningún caso* (ni siquiera en el suyo) porque '2+2 = 4' *es verdadero*, por eso nuestra obligación es corregir al niño. El niño cree que '2+2 = 4' no es verdadero y eso hace que su actitud de rechazo sea *racional* e *inteligible* para nosotros (aceptar '2+2 = 4' creyendo que no es verdadero habría sido irracional), pero su creencia *no hace que sea correcto* rechazar '2+2 = 4'. Otra prueba de que el objeto del rechazo es la falsedad y el de la aceptación la verdad (y no lo que *nos parece* falso o verdadero) es que, si tenemos motivos para dudar de la fiabilidad de nuestras facultades cognitivas (por hallarnos bajo los efectos de una droga –que no estimule nuestra disposición a aceptar o rechazar creencias– o en un entorno cuyas propiedades y leyes desconocemos totalmente), nos mostraremos reticentes a la hora de aceptar lo que *nos parece* verdadero o de rechazar lo que *no nos parece* verdadero, ya que podríamos estar equivocados. Esto muestra que nuestros objetivos al aceptar o rechazar algo son la verdad y la no verdad sin más.

Por todos estos motivos la primera respuesta de Priest a (B) no resulta satisfactoria, el lógico clásico puede insistir aún en que lo no verdadero (o lo falso) es el objetivo último del rechazo racional. Cuando algo no es verdadero (o cuando es falso), debe ser rechazado independientemente de que sea además no-verdadero (si es que el añadido realmente "añade" o cambia algo). En presencia de una dialetheia, p, el dialetheista *sólo* extrae las consecuencias que se siguen de su verdad: "p es aceptable", pero no las que se siguen de su no-verdad (o falsedad): "p es rechazable", p sería, pues, *aceptable y rechazable*, sin embargo nadie puede aceptar y rechazar algo al mismo tiempo. Frente a estas afirmaciones Priest tiene aún objeciones y observaciones que hacer con respecto a (B).

Según Priest, la verdad es *superior* a la falsedad en cierto sentido. Desde una perspectiva epistemológica, la verdad se halla por encima de la falsedad, ya que, por su naturaleza, la verdad es el objetivo último de los procesos cognitivos. Es constitutivo de la verdad ser "lo que uno debe aceptar". La falsedad, en cambio, no es sino la verdad de la negación, no tiene fuerza epistémica independiente. No es necesario, pues, rechazar la verdad de p simplemente porque $\neg p$ sea verdadera. Como la falsedad de una dialetheia, p, no elimina su verdad, debemos aceptar p y no rechazarla.[85]

[85] "If there is sufficient evidence that something is true, one ought, rationally, to accept it. [...] It is natural to suppose that there is a dual principle here: if there is sufficient

Este argumento es oscuro y no es obvio que consiga establecer lo que pretende. Es cierto que, a diferencia de la falsedad, la verdad tiene un valor y una relevancia especiales tanto *para nosotros* como *para nuestros procesos cognitivos*. Sólo podemos conocer *lo verdadero*, por consiguiente el fin último de cualquier proceso cognitivo (i.e., de cualquier proceso encaminado a obtener *conocimiento*) ha de ser forzosamente la verdad. Esto es así incluso cuando nos interesamos por la falsedad. Cuando nos preguntamos si *p* es falso (o si debemos rechazar *p*), lo que buscamos es *conocimiento* y, por tanto, lo que nos preguntamos es si *es verdad* que *p* es falso (o rechazable). Si quisiésemos *mentir*, la meta de nuestros procesos epistémicos *seguiría siendo la verdad* porque, para mentir "con éxito", deberíamos *saber* que lo que decimos es falso y por tanto deberíamos saber qué oraciones de la forma F(p) *son verdaderas*. Si no nos preocupase el conocimiento, *también* sería la verdad la meta de nuestros procesos cognitivos, aunque no *la nuestra*. Nuestro objetivo (no el de nuestros procesos cognitivos) sería entonces la *ignorancia*, suspender el ejercicio de nuestras facultades cognitivas. Consiguientemente, la falsedad o la no-verdad no podrían ser en ningún caso el fin de nuestros procesos cognitivos. En la medida en que

evidence that something is false, one ought, rationally, to reject it. If, therefore, there is strong evidence that contradictions, α and $\neg\alpha$, are both true, there is evidence that both are also false. One ought, then, to reject both." (Priest 1998b, p. 421) Sin embargo, Priest responde a continuación: "No. In the appropriate sense, truth trumps falsity. Truth is by its nature the aim of cognitive processes, such as belief. [...] It is constitutive to truth that that is what one ought to accept. Falsity, by contrast, is merely truth of negation. It has no independent epistemological force. One should not necessarily, therefore, reject something simply because its negation turns out to be true." (Ibíd., p. 421).

tengamos un fin cognitivo, éste ha de ser siempre la (búsqueda de la) verdad. En este sentido, "*truth trumps falsity*", como dice Priest.

Ahora bien, ¿qué conclusiones podemos extraer lícitamente de este hecho? Priest parece argumentar del siguiente modo: Cuando algo es sólo falso o no verdadero debe ser rechazado, sin embargo cuando algo es verdadero debe ser aceptado aunque sea también falso o no verdadero. La obligación de aceptar la verdad anula la obligación de rechazar la falsedad (o lo no verdadero) en el caso de las dialetheias. Esto es así porque la verdad es más importante que la falsedad, es el fin natural de nuestros procesos cognitivos, mientras que la falsedad es "sólo" la verdad de la negación.

Este argumento es insatisfactorio por varios motivos. Es insatisfactorio, en primer lugar, porque el hecho de que la verdad sea más importante que la falsedad *para nosotros*, no significa que, *en sí misma*, la verdad sea más importante o tenga más peso que la falsedad. De hecho, ni siquiera está claro que la verdad sea más importante que la falsedad para determinar nuestros fines cognitivos. Después de todo, no podríamos señalar *qué constituye* el fin último de nuestros procesos cognitivos (lo verdadero) si no pudiésemos señalar *aquello que no constituye* dicho fin (lo no verdadero o falso). En segundo lugar, este argumento es insatisfactorio porque no consigue establecer que las obligaciones epistémicas *objetivas* que impone la verdad de *p* ("debemos aceptar *p*") anulen o tengan más peso que las obligaciones *objetivas* que impone la no-verdad o falsedad de *p* ("debemos rechazar *p*"). Y, en tercer y último lugar, el argumento es insatisfactorio porque el hecho de que *definamos* la falsedad de *p* como "verdad de la negación de *p*" responde *a nuestros intereses, no a la naturaleza* de la falsedad. La falsedad de *p* no tiene por qué definirse como la verdad de la negación de *p*. Podríamos definir la verdad de *p* como la falsedad de la negación de *p*: $V(\underline{p})$ sii $F(\underline{\neg p})$ o, equivalentemente –si no somos dialetheistas–, como la negación de la falsedad de *p*: $V(\underline{p})$ sii $\neg F(\underline{p})$.

Frente a una oración, *p*, o a cualquier otra entidad que represente hechos del mundo (una proposición, una creencia, un enunciado, etc.), tanto Priest como el lógico clásico están de acuerdo en que lo que *p* dice (o expresa, o representa, o describe, etc.) *se da* o *no se da*. Es importante resaltar que este hecho es *independiente* de nuestra voluntad, intereses o fines cognitivos y que viene determinado por la naturaleza *representativa* de *p*. No tiene sentido decir que el "darse" de *p* es más *importante* que su "no darse". La palabra 'importante' está fuera de lugar en este contexto, hablamos de hechos, no del valor que dichos hechos tienen para nosotros o para nuestros fines, y es un hecho (para todo *p*) que *lo que dice p se da* o *no se da*. En el primer caso decimos que *p* es verdadera, en el segundo que no lo es (o que es falsa). Tal vez la verdad sea más importante *para nosotros* que la no-verdad en la medida en que es nuestro fin cognitivo, pero la verdad no es *en sí misma* más importante que la no-verdad o falsedad y, por consiguiente, lo que se sigue de la verdad no es *en sí mismo* más importante que lo que se sigue de la no-verdad. Priest podría replicar que las obligaciones *cognitivas* que se siguen de la verdad de un enunciado (su aceptación), *sí* son más importantes que las que se siguen de su no-verdad (su rechazo). Pero esto es erróneo. Las obligaciones cognitivas que se siguen en uno y otro caso son *igual* de importantes y no sólo en sí mismas sino también *para nosotros*. Veamos por qué.

Decir que "el fin de nuestros procesos cognitivos es la verdad", es *excluir* que dicho fin sea la no-verdad o la falsedad. Ahora bien, esta afirmación sólo tiene sentido si *hay algo que excluir*, es decir, si además de oraciones (proposiciones, etc.) verdaderas, hay también oraciones que *no son* verdaderas (no verdaderas significa aquí *sólo* no-verdaderas o *sólo* falsas). Si no hubiese oraciones (sólo) falsas o no-verdaderas, entonces *todas* las oraciones *serían verdaderas*, ya que, tanto para Priest como para el lógico clásico, lo que no es falso (o no-verdadero) es verdadero y viceversa: $\neg F(\underline{p}) \rightarrow V(\underline{p})$ y $\neg V(\underline{p}) \rightarrow F(\underline{p})$ (donde '$\rightarrow$' satisface *Modus Ponens*).[86] Ahora bien, si *todo* fuese verdadero, no tendría sentido decir que la verdad es el fin de nuestros procesos cognitivos, como tampoco tiene sentido decir que el fin de una actividad es "seleccionar objetos idénticos a sí mismos". Si *todos* los objetos son idénticos a sí mismos –olvidémonos por un momento de las paradojas–, entonces no hay nada que seleccionar. Si todas las oraciones fuesen verdaderas, entonces no habría ningún fin que perseguir, buscar la verdad sería como viajar justo a donde *ya estamos* sin dar la vuelta al mundo. Pero las consecuencias son aún peores. Si *todo* fuese verdadero, entonces *todo* sería *no-*

[86] Priest justifica esta tesis apelando a la existencia de "hechos negativos" (véase, el final de 5.2.2.2): Cuando una oración, *p*, no es falsa (verdadera) por no existir ningún hecho en el mundo que la haga falsa (verdadera), entonces *p* es verdadera (falsa) en virtud de un hecho "negativo", i.e., el hecho de que "no existe ningún hecho ("positivo") que haga falsa (verdadera) a *p*". Los hechos negativos *sólo* pueden hacer falsas o verdaderas a oraciones que no sean dialetheias (oraciones *sólo* verdaderas o *sólo* no-verdaderas que son precisamente las que consideramos aquí), es, pues, pertinente apelar a ellos en este contexto.

verdadero y *falso* (*trivial*), ya que las oraciones $\forall x \neg V(x)$ y $\forall x F(x)$ así como todas sus ejemplificaciones *serían verdaderas*. Así pues, las oraciones verdaderas, no-verdaderas y falsas serían indistinguibles y decir que el fin de nuestros procesos cognitivos es la verdad sería equivalente a decir que dicho fin es la no-verdad o lo falso. Así pues, no podemos establecer cuál es el fin de nuestros procesos cognitivos si no hay oraciones verdaderas *y oraciones no-verdaderas*. El sentido de esta discusión es establecer que lo no verdadero es *tan* importante y necesario para determinar y perseguir nuestros fines cognitivos *como* lo verdadero. No hay razón alguna, por tanto, para pensar que las obligaciones cognitivas que nos impone la verdad sean *más* importantes que las que nos impone lo falso o lo no-verdadero. Perseguir la verdad implica *discernir* cuándo algo *es* verdadero y cuándo *no lo es*. Lo verdadero *debe ser aceptado*, lo no verdadero *debe ser rechazado*. Si no aceptamos lo verdadero y rechazamos lo que no es verdadero nuestro propósito de buscar la verdad fracasa. Ambas obligaciones están en pie de igualdad, esto está claro cuando tropezamos con oraciones que son sólo verdaderas o sólo no verdaderas, pero también debería estarlo cuando tropezamos con dialetheias. Si existen oraciones que son verdaderas y no verdaderas estamos racionalmente obligados *a aceptarlas y a rechazarlas*. Si decidimos aceptar una dialetheia y no rechazarla sólo podemos hacerlo sobre la base de criterios arbitrarios, no de criterios racionales.[87] Lo

[87] Priest está de acuerdo en que, de haber enunciados racionalmente aceptables y rechazables, nuestra obligación de aceptarlos no anularía nuestra obligación de rechazarlos. En una discusión relativa a la cuasi-validez de la ley de silogismo disyuntivo (recogida en un principio que guarda relación con la máxima metodológica

verdadero *no es* más importante que lo no verdadero (o lo falso) y la falsedad *no es* "meramente" la verdad de la negación (o, al menos, no más de lo que la verdad pueda ser la falsedad de la negación).

Priest contempla a veces la posibilidad de que las dialetheias fuesen aceptables y rechazables y concluye que esto no impediría que fuesen verdaderas y falsas (o incluso verdaderas y no-verdaderas, como ocurre con $\underline{\mu}$). El hecho de que la aceptación y el rechazo de p sean psicológicamente incompatibles no establece que la verdad de p (el *telos* de la aceptación) sea incompatible con su no-verdad (el *telos* del rechazo).[88] Priest señala además que en ciertas situaciones, por ejemplo en contextos legales pero también en contextos normativos de otro tipo como la práctica de la epistemología, nos percatamos de que algunas de nuestras obligaciones son contradictorias y no pueden ser satisfechas conjuntamente. En esas circunstancias tenemos obligaciones incompatibles. Más que una posibilidad, esto es un hecho contrastable empíricamente, ahora bien, ¿qué moraleja hemos de extraer de este hecho? El dialetheista pensará que la existencia de obligaciones incompatibles (como por ejemplo las de aceptar y rechazar $\underline{\mu}$ = '$\neg V(\underline{\mu})$') es sencillamente inevitable, la inconsistencia es un rasgo de la

---

de la nota 56), Priest dice: "Some confusión may arise from the thought that something's being rationally acceptable (as well as rejectable) 'cancels out' its rational rejectability. This is just a confusion. If something is rationally acceptable and rejectable, it is still rationally rejectable. Any consequences that this fact has still, therefore, stand." (*IC*, p. 142).

[88] "If two kinds of actions are incompatible, why should one suppose that their *teloi* are also incompatible? Notoriously there is more than one way to skin a cat: we may, in fact, achieve not merely compatible, but even the same end by very different, and incompatible means. One may strive for peace by arming to teeth as a deterrent, or by totally disarming, and persuading others that one is no threat." (Priest 1993, p. 40).

racionalidad y por eso a veces estamos obligados a hacer *lo imposible*, es decir estamos *condenados* a no satisfacer alguna de nuestras obligaciones.[89] El lógico consistente, por su parte, pensará que debemos rechazar, restringir o reformar aquellos principios que nos conducen a la aceptación de obligaciones incompatibles, no hacerlo *sería irracional.* Sin embargo, la situación no es ni mucho menos de tablas, como pretende Priest. Si estamos condenados a no satisfacer una de nuestras obligaciones racionales, aceptar $\underline{\mu}$ o rechazar $\underline{\mu}$, ¿qué razones tenemos para elegir una de las dos antes que la otra? Priest elige aceptar $\underline{\mu}$, pero podríamos optar perfectamente por rechazar $\underline{\mu}$ junto a cualquier otra dialetheia. Así pues, Priest debe justificar su elección.

El dialetheista podría (previsiblemente) maniobrar de la siguiente manera: "Si hubiese dialetheias, entonces (como dice mi oponente) tendríamos la obligación racional de aceptarlas y rechazarlas. Como no podemos satisfacer ambas obligaciones porque son incompatibles, *debemos optar por una*. "*C'est la vie*". Una buena razón para aceptar las dialetheias, en vez de rechazarlas, es que *no existe ninguna teoría consistente satisfactoria que elimine las paradojas*. Las paradojas parecen establecer que un mismo enunciado, *p*, es verdadero y falso (o no verdadero), por tanto es mejor (más satisfactorio) aceptar que *p* es verdadero y falso que rechazarlo. Por otra parte, estos dilemas son

---

[89] "Suppose that untruth is the telos of rejection. If $\underline{\lambda}$ [$\lambda \leftrightarrow \neg V(\underline{\lambda})$] is both true and untrue, and we have evidence of this fact, should one accept or reject it? [...] we can be put in a bind where we are obliged to do the impossible. [...] Maybe rational obligation can produce similar binds too; rationally, we are damned if we do and damned if we don't. $\underline{\lambda}$, then, should be both accepted and rejected. This is impossible. *C'est la vie*." (Priest 1993, pp. 40-1, véase también *IC*, capítulo 13.)

precisamente el tipo de problemas con los que previsiblemente uno tropezaría *si hubiese dialetheias*, pero es que, además, *éstos son los problemas con los que, de hecho, tropezamos habitualmente, ergo debe haber dialetheias*."

Después de recorrer un largo camino, hemos vuelto a tropezar con la misma piedra, "no existe una solución consistente satisfactoria a las paradojas". Este hecho se utiliza ahora para justificar una elección no arbitraria entre dos obligaciones racionales incompatibles: Aceptar y rechazar las dialetheias. Decidimos aceptar y no rechazar las dialetheias (pese a estar obligados a rechazarlas) porque "no existe una solución consistente satisfactoria a las paradojas". La pregunta clave es, pues, evidente: ¿Es el dialetheismo una solución inconsistente *satisfactoria* a las paradojas?

5.3.5 Incoherencias y paradojas reforzadas propias del dialetheismo

Uno podría argumentar que el dialetheismo es una solución insatisfactoria a las paradojas por motivos muy similares a los que aduce el dialetheista para desacreditar cualquier solución consistente a las mismas. El dialetheismo restringe la validez de principios generales de inferencia intuitivos y modifica el significado habitual de términos básicos de nuestro lenguaje y todo ello recurriendo a menudo a distinciones y justificaciones *ad hoc*. Sin embargo el dialetheista reclama para su teoría un mérito que niega a sus oponentes. En cierto sentido, el dialetheismo "*deshace*" las paradojas. Las alternativas al dialetheismo prometen solucionar las paradojas consistentemente, pero sólo consiguen generar una teoría consistente para una fracción de nuestro lenguaje que ni siquiera incluye los medios necesarios para expresar la propia teoría. Esto supone un problema porque la teoría contiene a su vez conceptos que dan lugar a nuevas paradojas que han de ser solucionadas apelando a una (meta-)teoría distinta. Se genera así una jerarquía de teorías, o incluso de lenguajes, que no puede ser descrita mediante una teoría general consistente. Al aceptar inconsistencias el dialetheista puede ofrecer, en cambio, una teoría *general* sin recurrir a jerarquías de ningún tipo, lo que se adecua mejor a nuestra visión de los lenguajes naturales. Según el dialetheista, *algunas* proposiciones son verdaderas y falsas, por tanto no pasa nada si nuestra teoría establece y acepta la verdad de *algunas* contradicciones. Si llegamos a comprender que aceptar *algunas* contradicciones *no supone aceptar un absurdo*, entonces la inferencia de una contradicción en ciertos contextos *dejará de ser algo paradójico* y pasará a ser una consecuencia natural de ciertas premisas y principios teóricos perfectamente válidos.

Pero el planteamiento del dialetheista tropieza aquí con un problema. Hay que decidir *qué contradicciones son* aceptables y *cuáles no lo son* y ésta no es una tarea fácil. Vimos en 5.2.2.1 que *no todas* las contradicciones pueden ser verdaderas, de lo contrario el dialetheismo sería una doctrina *trivial*, pero también vimos en 5.3.1 que carecemos de algoritmos o criterios sintácticos para decidir qué contradicciones (u oraciones en general) debemos rechazar. Frente a cualquier contradicción el dialetheista debe decidir si es o no una dialetheia sopesando las pruebas de que dispone en uno y otro sentido. Este proceso de evaluación será siempre falible y revisable y no puede presuponer en ningún momento la verdad del dialetheismo. Priest le ha exigido al lógico consistente que no incurra en una petición de principio, es justo que el lógico consistente pida ahora lo mismo. El dialetheista no puede decir: "Como algunas contradicciones son verdaderas (así lo demuestra el fracaso de las soluciones consistentes a las paradojas), no pasa nada si inferimos $p \wedge \neg p$ en nuestra teoría. Siempre que esto ocurra, p es una dialetheia". Este argumento es inválido, no se puede justificar la verdad de una contradicción apelando exclusivamente a la verdad del dialetheismo, necesitamos siempre razones independientes, de lo contrario el dialetheismo sería una teoría inatacable *por definición*, no por méritos objetivos.

Si la teoría de Priest genera paradojas que no puede resolver, es decir, si permite inferir resultados absurdos o contradicciones inaceptables incluso para un dialetheista, entonces será tan insatisfactoria como el resto de teorías examinadas en el capítulo cuatro y, seguramente, por los mismos motivos. El propio Priest concede este punto:

> "Suppose that it turned out, in defending the inconsistent view, that it had to be shorn up in the same methodologically unsatisfactory ways as extant consistent accounts –for example, to avoid strengthened paradoxes– until it was just as complex and contrived. It would then cease to be rational to accept it. The fact that one can accept some contradictions would do nothing to help the matter."[90]

[90] Priest 1998b, p. 423.

Es importante insistir en esta idea. Si para evitar la inferencia de absurdos o de contradicciones que no son dialetheias Priest se ve forzado a introducir arbitrariamente restricciones expresivas en su teoría (i.e., restricciones destinadas a impedir la expresión de paradojas no deseadas), entonces su situación será idéntica a la del lógico consistente (LC), o quizá peor. Al igual que LC, Priest habrá restringido la validez general de varios principios de inferencia, habrá introducido distinciones *ad hoc* y no habrá conseguido construir una teoría *general* coherente en la que *todas* las contradicciones inferidas sean *aceptables* (i.e., dialetheias). Y si algunas contradicciones continúan siendo paradójicas, será preciso una nueva teoría que las resuelva, lo que abre las puertas a la formación de una jerarquía de teorías. Si Priest se viese forzado a aceptar todo esto, su teoría incumpliría lo prometido y sería consiguientemente tan insatisfactoria como sus rivales. Así pues, ¿es el dialetheismo una teoría satisfactoria?

Para atacar la teoría de Priest sería preciso construir una *reductio* en la que mostrásemos que el dialetheismo permite inferir proposiciones absurdas no sólo desde la perspectiva del lógico clásico sino también del dialetheista. La *reductio* que buscamos no tiene por qué basarse en una contradicción, podemos buscar sencillamente la inferencia de un enunciado absurdo o inaceptable para el dialetheista, por ejemplo "0 = 1" o "Todo es verdadero".[91] Eso sí, para que la *reductio* tenga fuerza sólo

[91] Esta estrategia es seguida por varios críticos: "[C]ontradictions are not the only pivot for a reductio proof. Showing that a theory leads to triviality, in the shape of the proposition that everything is true, would be a knock-down argument against it." Smiley 1993, p. 27 y siguientes. Bromand (2000, p. 746) también sigue esta estrategia, a partir

puede apoyarse en principios de inferencia válidos para el dialetheista. Si tropezamos con una proposición absurda de estas características, entonces deberemos rechazar el dialetheismo.

Hemos señalado ya que Priest no ignora la existencia de contradicciones internas en su teoría. El dialetheista no sólo acepta la verdad y falsedad de enunciados como $\underline{\mu}$ = '$\neg$V($\underline{\mu}$)', también acepta la verdad y falsedad de algunos de los enunciados que emplea para describir su propia teoría, por ejemplo: "Hay oraciones que son verdaderas y no-verdaderas". Tanto $\exists x(V(x) \wedge \neg V(x))$ como su negación son dialetheias. Apuntamos este hecho al final de 5.2.2.2 y discutimos sus consecuencias en 5.3.4. Priest reconoce la existencia de varias contradicciones teóricas internas (más depuradas en 1987 que en 1979, véase la nota 81) y justifica este hecho diciendo que si el dialetheismo ha de aceptar contradicciones, éstas han de afectar al conjunto de la teoría. No se trata de construir una teoría inconsistente mediante un discurso metateórico consistente sino de construir una única teoría inconsistente que sea capaz de dar cuenta de sí misma. Esta teoría puede contener y contiene contradicciones "teóricas", es decir, inconsistencias que afectan a tesis generales sobre la verdad y la falsedad de sus enunciados.[92] Pero estas

de un argumento que establece 0 = 1, infiere que todo es verdadero ya que todos los valores de verdad dialetheistas son idénticos a {1}, i.e., "verdadero". Priest sí concede validez a una *reductio* de este tipo (aunque no acepta las *reductio* que ofrecen sus oponentes). Véase Priest 1993.

[92] Véase sobre todo la cita de la nota 80. Esta idea se mantiene desde los principios del dialetheismo: "[O]nce we have given up demanding that the object theory be consistent, there is no reason to demand that the metatheory be consistent. Indeed, this is forced on us if we wish to give a coherent account of paradoxicality. Any object theory

contradicciones no son arbitrarias. Todas ellas derivan de razonamientos o inferencias cuyo punto de arranque es una paradoja de inclusión. Después de todo, *sólo* podemos probar que (1) $\exists x(V(x) \wedge \neg V(x))$ es una dialetheia si $\underline{\mu}$ (i.e., '$\neg V(\underline{\mu})$') es una dialetheia. (1) es una dialetheia *porque* μ lo es. La aceptación del carácter paradójico de (1) es para el dialetheista una muestra de la *coherencia* de su teoría.[93]

Aunque Priest tuviese algo de razón en este punto es importante señalar que sólo tendría razón "a medias". En el fondo *sí* existe un problema de *coherencia expositiva*.[94] Si (1) es una dialetheia y, por tanto, falsa, entonces la negación de (1), i.e., (2) $\neg\exists x(V(x) \wedge \neg V(x))$, es verdadera por definición, pero lo que (2) nos dice es que "*No hay* oraciones que sean verdaderas y no verdaderas". Al exponer su teoría el dialetheista siempre hace hincapié en (1) y no en (2) pese a ser una tesis igualmente válida desde el punto de vista de su teoría. ¿Cuándo hemos de creer al dialetheista, cuando nos dice que *hay* oraciones verdaderas y no verdaderas o cuando nos dice que *ninguna* oración es verdadera y no verdadera? ¿Qué nos está diciendo? Priest sólo subraya aquellas consecuencias de su teoría que le interesan, las que niegan la posibilidad de una dialetheia son ignoradas pese a ser igualmente válidas. Por otra parte, el dialetheista no es completamente *consecuente* con lo que dice,

---

inconsistency $A \wedge \neg A$ is transformed simply by the *T*-scheme into a metatheoretical contradiction ⌈A is true and A is not true⌉." (Priest 1979, V, p. 239).

[93] Hablando de (1) y de su negación Priest dice: "in virtue of the fact that the contradictory claims are semantical, and that a self referential construction (the liar paradox [μ]) was necessary to prove one of them [(1)], this is exactly where I have urged that contradictions should be expected to turn up. In a sense, therefore my position is quite [...] self-coherent." (*IC*, p. 91).

como ya señalamos en 5.3.4. Si (1) es una dialetheia y una dialetheia es falsa (además de verdadera), entonces (1) debería ser rechazada (además de aceptada). Pero Priest sólo atiende algunas de las obligaciones que le impone el dialetheismo (aceptar dialetheias) y desatiende sistemáticamente otras que también se siguen de su teoría (rechazar dialetheias).

Ninguna tesis dialetheista justifica el trato asimétrico que reciben las oraciones demostrablemente verdaderas y demostrablemente falsas del dialetheismo. Priest sugiere que la verdad es "más importante" que la falsedad desde el punto de vista de nuestros procesos cognitivos pero vimos en 5.3.4 que este argumento no es convincente. También consideramos entonces una posible respuesta por parte del dialetheista para justificar la asimetría: "Aceptamos las dialetheias (y primamos su verdad en la exposición de nuestra teoría) porque no hay soluciones consistentes satisfactorias a las paradojas, lo que parece favorecer la existencia de dialetheias." Está claro, no obstante, que Priest no podría defenderse de este modo sin incurrir en una petición de principio. Para que esta réplica fuese efectiva debería demostrar que *no es posible* solucionar de forma consistente las paradojas. Priest ha mostrado, a lo sumo, que ninguna de las soluciones *existentes* es satisfactoria, pero esto no excluye la posibilidad de encontrar una que sí lo sea en el futuro.[95]

---

[94] Esta objeción se debe en lo fundamental a Simmons (1993, pp. 80-2).

[95] La imposibilidad de descartar una solución consistente se reconoce implícitamente en la siguiente cita: "Naturally, it may happen that someone, a hundred years hence, will come up with a consistent account of truth with none of these problems, in which case, what it is rational to believe may well change. But that is neither here nor there. Rational belief about anything is a fallible matter. It is a mistake to believe where the evidence

Ciertamente, podría no haber ninguna, pero eso no demostraría que la teoría de Priest es satisfactoria. Priest aún está obligado a demostrar que su teoría está libre de paradojas, de lo contrario podría no haber soluciones satisfactorias a las paradojas (ni consistentes ni inconsistentes). El dialetheista no puede presuponer la verdad del dialetheismo para solucionar sus problemas, debe ofrecer siempre argumentos independientes.

Varios autores han señalado que el dialetheismo, lejos de eliminar las paradojas, genera versiones reforzadas de las mismas al igual que sus competidores. Sus principios teóricos conducirían a la inferencia de absurdos tanto para el lógico clásico como para el dialetheista, absurdos que involucrarían además conceptos básicos empleados por el dialetheista en la exposición de su teoría. Los mismos críticos han señalado que en su intento por evitar paradojas reforzadas, Priest ha restringido seriamente la capacidad expresiva del lenguaje de su teoría. Nos centraremos aquí en dos textos: Smiley 1993 y Bromand 2002.

Smiley (1993, pp. 27-32) ofrece tres reducciones al absurdo de la teoría de Priest. La primera y la tercera parten de supuestos más claros que la segunda, por eso nos centraremos en ellas.[96] La primera (ibíd., pp.

---

does not point; but it is equally a mistake not to believe where the evidence points." (Priest 1998b, p. 421). La última afirmación no anula la fuerza de la primera.

[96] La segunda involucra las nociones de demostración [*proof*] y refutación [*disproof*]. Para el lógico clásico refutar p y demostrar $\neg p$ son lo mismo, para el dialetheista, presumiblemente, no: Si δ es una dialetheia (δ sii $\neg\delta$), entonces demostrar δ equivale a demostrar $\neg\delta$, pero no a refutar δ (refutar δ implicaría rechazar δ). Razonando a partir de la oración A = 'A es refutable', Smiley (ibíd., pp. 29-30) demuestra B (una fórmula arbitraria) apelando a principios dialetheistas, y concluye que el dialetheismo desemboca en el trivialismo. Priest (1993, pp. 47-9) rechaza esta *reductio* atacando la noción de implicación que aparece en algunos de los principios utilizados por Smiley en

27-9) atañe a la validez de la teoría intuitiva o "naif" de conjuntos que Priest acepta y cuyo principio de extensionalidad afirma en parte lo siguiente: "$\{x \mid \alpha\} = \{x \mid \beta\}$ es falso sii $\alpha(x/t)$ es verdadero y $\beta(x/t)$ es falso (o viceversa) para algún término cerrado, t." (*IC*, p. 184, α y β son fórmulas abiertas con una sola variable libre, x, y $\alpha(x/t)$ es el resultado de sustituir x por t en α). En la teoría "naif" de Priest $R = \{x \mid x \notin x\}$ es un conjunto, por lo tanto podemos interpretar α y β como $x \notin x$. La paradoja de Russell ($R \in R$ sii $R \notin R$) nos permite inferir entonces: $\alpha(x/R)$ es verdadero y $\beta(x/R)$ es falso, de donde se sigue, por el principio de extensionalidad, $R \neq R$. Consideremos ahora la fórmula F(x) (F(x) es verdadera sii $x = R$) y las siguientes oraciones:

(*) Hay un F (**) Hay dos Fs

Smiley formaliza (*) y (**) de la siguiente manera:

(*) $\exists y \forall x(\mathrm{F}(x) \leftrightarrow x = y)$ (**) $\exists y \exists z \forall x(y \neq z \wedge (\mathrm{F}(x) \leftrightarrow x = y \vee x = z))$

Ambas oraciones son demostrables apelando a principios válidos para el dialetheista: (*) se obtiene al introducir '$\exists$' en el siguiente teorema: $\forall x(\mathrm{F}(x) \leftrightarrow x = R)$; y (**) se obtiene de modo similar a partir del teorema: $\forall x(R \neq R \wedge (\mathrm{F}(x) \leftrightarrow x = R \vee x = R))$. Dado que (*) y (**) son ambas verdaderas, concluimos que el número de Fs es 1 y 2, *de donde inferimos un absurdo*, 1 = 2, apelando al principio de la aritmética que relaciona, según Smiley, el uso de numerales como adjetivos y su uso como sustantivos: "El número de Fs es *n* si hay *n* Fs".

su demostración. Según Priest, Smiley aplica la ley de modus ponens sobre condicionales que no la satisfacen. La demostración es, pues, inválida para un dialetheista.

Priest (1993, pp. 45-7) rechaza esta *reductio*. Aunque el argumento fuese correcto, no se seguiría obviamente de 1 = 2 que *todo* fuese verdad (el trivialismo), si bien es cierto que 1 = 2 resulta por sí solo bastante nefasto. Por fortuna, no tenemos que aceptar esta ecuación. El argumento de Smiley es incorrecto porque no podemos inferir "1 = 2" a partir de "el número de Fs es 1 y 2". El principio que relaciona el uso de numerales como adjetivos con su uso como sustantivos fue determinado por Frege y Cantor: "El conjunto de los Fs pertenece al número n sii hay n Fs".[97] Para establecer 1 = 2, Smiley necesita demostrar la validez del siguiente principio: "Si n y m son números naturales, entonces $\forall x(x \in n \cap m \rightarrow n = m)$". Este principio nos permitiría inferir 1 = 2 a partir de $R \in 1 \cap 2$, sin embargo, este principio *no es válido en general* para el dialetheista. Es cierto que el principio se cumple cuando contamos (los elementos de) conjuntos *consistentes*, pero no tenemos ninguna razón para pensar que también deba cumplirse cuando contamos conjuntos inconsistentes. Por otra parte, *nada* de lo que dice el dialetheismo *impide* que el principio mencionado se cumpla en el caso de conjuntos consistentes y nada de lo que dicen nuestras prácticas habituales nos

[97] Desde esta perspectiva, un número n se concibe como un conjunto cuyos miembros son *todos* los conjuntos *con n elementos*. Para determinar que un conjunto A tiene n elementos debemos establecer una relación que *empareje uno a uno todos* los elementos de S con *todos* los elementos de un conjunto B con n elementos, i.e., un B tal que $B \in n$. Para ello elegimos como B a un "representante" de los elementos de n con el que comparar otros conjuntos. El representante de 0 puede ser $0_r = \{x\colon V(x) \text{ sii } \forall y V(y)\} = \varnothing$ (el dialetheista no puede elegir $\{x\colon x \neq x\}$ como 0_r ya que no es vacío: $R \in \{x\colon x \neq x\}$); el representante de 1 es $1_r = \{0_r\}$ y en general $n_r = \{0_r, ..., n-1_r\}$. La defensa y exposición más celebrada de esta manera de concebir la relación entre numerales usados como adjetivos y como sustantivos (y la consiguiente justificación) se encuentra en Frege 1884.

indica cómo contar conjuntos inconsistentes. Un rasgo característico de dichos conjuntos, por oposición a los consistentes, podría radicar precisamente en el hecho de que *no tuviesen un único* número de elementos.[98] La carga de la prueba cae, pues, en el lado del matemático clásico. Según Priest, un argumento a favor de $\forall x(x \in n \cap m \rightarrow n = m)$ involucraría probablemente el principio de transitividad de la implicación (TRI) que, como ya señalamos en 5.2.3.1, no tiene validez general para el dialetheista (TRI es cuasi-válido, i.e., válido sólo en contextos consistentes). Tomemos dos conjuntos x y z de cardinalidad n y m respectivamente. Para establecer $n = m$ a partir de $x \in n \cap m$ seguramente necesitaríamos establecer un principio del tipo: "si existe una correlación uno a uno entre x e y y existe una correlación uno a uno entre y y z, entonces existe una correlación uno a uno entre x y z". Sin embargo la aplicación de este principio involucraría en la práctica el uso de TRI, un principio que sólo es cuasi-válido.

La última *reductio* considerada por Smiley, la más simple e incisiva, descansa sobre una paradoja semántica reforzada. Consideremos la siguiente oración: $\underline{\delta}$ = '$\underline{\delta}$ es *sólo* falsa' (δ sii $v(\delta) = \{0\}$). Priest establece (véase 5.2.3.1) que una oración $\underline{p}$ es verdadera sii $1 \in v(p)$ y que $\underline{p}$ es falsa sii $0 \in v(p)$, de este modo el esquema tarskiano (PEV) se

[98] "Does the fact that a set can have more than one number not play havoc with our normal practice of counting? No! We have no normal practice of counting inconsistent collections. Our normal practice is of counting quite consistent collections, such as the number of marbles in a tin; and dialetheism gives no reason at all to suppose that the numbers of such collections are not unique. Uniqueness is violated only in the case in which we are counting inconsistent totalities. And if you will try to count these, what do you expect?!" (Priest 1993, p. 46).

puede expresar mediante: $1 \in v(p)$ sii p. Priest también establece que, para toda oración p, *sólo* existen tres posibilidades evaluativas (exhaustivas y *excluyentes entre sí*): o $\underline{p}$ es *sólo* verdadera: $v(p) = \{1\}$; o *sólo* falsa: $v(p) = \{0\}$; o verdadera *y* falsa: $v(p) = \{0, 1\}$. Siendo éste el caso, existen tres posibilidades con respecto a $\underline{\delta}$: O $v(\delta) = \{1\}$, o $v(\delta) = \{0\}$, o $v(\delta) = \{0, 1\}$. En el primer y tercer caso $\underline{\delta}$ es verdadera: $1 \in v(\delta)$, de donde inferimos, aplicando PEV, que $\underline{\delta}$ es sólo falsa: $v(\delta) = \{0\}$. En el segundo caso, $\underline{\delta}$ es sólo falsa y concluimos, aplicando PEV de nuevo, que es verdadera: $1 \in v(\delta)$. Así pues, tenemos en los tres casos: $v(\delta) = \{0\}$ y $1 \in v(\delta)$, por tanto $1 \in \{0\}$ o, en otras palabras, $0 = 1$. Esto significa que todos los valores semánticos que puede tomar una oración son iguales a $\{1\}$: $\{0\} = \{1\} = \{0, 1\}$, dado que *toda* oración tiene un valor de verdad, se cumple, para *cualquier* oración $\underline{p}$, que $\underline{p}$ es verdadera: $v(\underline{p}) = \{1\}$, y, por tanto, el dialetheismo es una teoría trivial.

Priest es consciente de este tipo de paradojas desde el principio, en 1979 (V.3, pp. 239-40) considera una muy parecida: 'Esta oración no es *sólo* verdadera', allí se limita a decir, una vez más, que esta oración es paradójica pero que, como el fin del dialetheismo no es erradicar las paradojas sino acomodarlas y no existen diferencias entre paradojas teóricas y metateóricas, la existencia de oraciones como ésta o como $\underline{\delta}$ no debería inquietarnos, ambas son *dialetheias*. Pero, ¿es verdad que estas oraciones son dialetheias? Es cierto que ambas son contradicciones *de inclusión* o paradojas de *autorreferencia*, es decir, el tipo de enunciados del que cabría esperar problemas según la teoría de Priest, sin embargo esto no determina nuestra actitud ante ellas o su evaluación semántica.

Priest reconoce que *no existen rasgos sintácticos* que determinen cuándo una oración es una dialetheia y cuándo no, o cuándo una contradicción es rechazable y cuándo no. El mero hecho de que estas oraciones compartan ciertos rasgos sintácticos y estructurales con otras oraciones que son dialetheias *no determina* que debamos considerarlas dialetheias (véase 5.3.1). Según Priest, hemos de examinar las pruebas de que disponemos a favor de la verdad o aceptabilidad de $\underline{\delta}$ o a favor de su falsedad o rechazo. Lo que el argumento anterior muestra es que no es posible construir una teoría dialetheista coherente con tres valores de verdad excluyentes ya que entonces podríamos mostrar a partir de $\underline{\delta}$ que todos ellos se reducen a uno: 'verdadero', desembocando en el trivialismo. Aceptar $\underline{\delta}$ como una dialetheia es aceptar todas las consecuencias que se siguen "dialetheísticamente" de su verdad y falsedad, ahora bien, lo que esas consecuencias dictan es que el dialetheismo es, en contra de lo que afirma, una doctrina trivial. Un dialetheista no puede acomodar en su teoría las contradicciones que se siguen de $\underline{\delta}$. Si $\underline{\delta}$ fuese una dialetheia, sería "verdadera y falsa" y "sólo falsa", así pues, decir que $\underline{\delta}$ es una dialetheia lejos de resolver el problema *lo genera*. Si Priest aceptase $\underline{\delta}$, su discurso sería, además de inconsistente, *incoherente* puesto que no lograría separar su doctrina del trivialismo, el motivo por el que rechazó ECQ.

La paradoja señalada por Smiley no coge desprevenido a Priest, quien ya considera $\underline{\delta}$ y sus consecuencias en Priest 1984b. La conclusión que extrae allá es que podría haber oraciones inconsistentes, a las que llama hiper-contradicciones, *con más de un valor de verdad dialetheista*, por ejemplo oraciones "verdaderas y falsas y sólo verdaderas".[99] En el artículo citado, Priest explora en qué medida deberíamos modificar LP, su lógica paraconsistente, para acomodar estas inconsistencias y evitar sus consecuencias. Priest reforma la semántica dialetheista para extender LP a una nueva lógica LP*. En la semántica de la lógica clásica, S_{LC}, los valores de verdad que puede tomar una oración son 2, $S_{LC} = \{0, 1\}$; en LP son $2^2 - 1$, $S_{LP} = \{\{0\}, \{1\}, \{0, 1\}\}$; en la extensión LP* de LP que buscamos serán $2^3 - 1$. Es decir, S_{LP*} incluirá a todos los *subconjuntos* de valores de S_{LP} excepto al conjunto vacío (i.e., todos los valores de S_{LP} junto a todas las combinaciones dobles de elementos de S_{LP} más el valor "ser sólo verdadero, sólo falso, y verdadero y falso". Sólo se excluye la posibilidad de que un enunciado no tenga *ningún* valor de verdad). El problema con esta solución es que *también* podemos construir hiper-contradicciones en LP* lo que nos fuerza a extender su semántica de nuevo para obtener una lógica LP** con $2^7 - 1$ valores de verdad. No hace falta decir que esta operación se repite una y otra vez debido a la presencia de hiper-contradicciones de rango superior hasta obtener una jerarquía de lógicas diferentes que no pueden ser abarcadas o expresadas en términos de una sola lógica *libre de hiper-contradicciones*. Esta solución no es sino una variante de las soluciones paramétricas que examinamos al final de 4.3.1 y resulta insatisfactoria por los motivos expuestos en 4.3.2.

[99] "[I]f there are sentences which are so twisted as to take impossible values, such as both true and false, might there not be sentences which are so contradictory as to take impossible values such as both true and false ($\{0, 1\}$) and true only ($\{1\}$)? Indeed there

No es de extrañar, pues, que a partir de 1987 Priest olvide este enfoque y vuelva a una semántica con sólo tres posibilidades evaluativas. Su respuesta a $\underline{\delta}$ en 1987 (véase 5.2.3.1) y en 1993 (pp. 50-1) busca evitar la paradoja haciendo que $\underline{\delta}$ sea *inexpresable* en LP. Con este fin, Priest *niega* que las evaluaciones tengan un carácter *funcional* en LP. Si evaluamos oraciones mediante funciones, entonces cada oración podrá recibir sólo un *único* valor de verdad, sin embargo $\underline{\delta}$ recibe más de uno, como acabamos de ver. Una manera de eludir el problema es hacer que las evaluaciones semánticas *sean relaciones, pero no funciones* (su presentación funcional en *IC* sería secundaria, sólo facilita la exposición, pero nada más), ya que algo puede mantener la misma relación con *más de un* objeto. Priest renunciaría, así, a hablar de "*valores*" de verdad (un vocablo que sugiere la evaluación de oraciones mediante funciones que asignan un *único valor* a cada uno de sus argumentos). Dada una relación evaluativa, v, y una oración p: p es verdadera sii $v(p, 1)$; p es falsa sii $v(p, 0)$; y p es verdadera y falsa sii $v(p, 1)$ y $v(p, 0)$. La única manera que tenemos ahora de expresar $\underline{\delta}$ (i.e., '*Esta oración es sólo falsa*') es la siguiente: δ sii $v(\delta, 0) \wedge \neg\, v(\delta, 1)$. A partir de nuestras reflexiones sobre $\underline{\delta}$ podemos inferir $v(\delta, 1) \wedge \neg\, v(\delta, 1)$, pero esta contradicción *es una dialetheia ordinaria* y *no una paradoja reforzada* para el dialetheista. Priest concluye que la objeción de Smiley se puede sortear apelando al carácter arbitrario de nuestra notación semántica. Existen formas de

---

are, as a simple application of the extended liar shows. Consider the sentence: this sentence is false only." (Priest 1984b, pp. 238-9).

representar la semántica de LP que son más coherentes que otras con las tesis semánticas del dialetheista y que son, por ello, preferibles.

¿Es ésta una respuesta satisfactoria? Es cierto que evitamos la paradoja reforzada cambiando nuestra notación semántica pero, ¿qué coste tiene esta maniobra? Smiley (op. cit., pp. 31-2) señala que el precio a pagar es alto y no incumbe sólo a las funciones que evalúan oraciones. Muchas son las funciones f (en matemáticas, en lógica y en otras disciplinas relacionadas) que definimos por casos según el patrón:

$f(x) = y$ si $P(x)$

$f(x) = z$ si $\neg P(x)$

(donde $y \neq z$). Un dialetheista acepta la posibilidad de que $P(x) \wedge \neg P(x)$ sea verdadera en algunos casos, lo que significa que $f(x) = y$ y $f(x) = z$ pero $y \neq z$. Ahora bien, *ninguna* función puede satisfacer esta condición, ya que, *por definición*, el rasgo *constitutivo* de toda función es la unicidad del valor asignado a cada uno de sus argumentos:

$\forall xyz((f(x) = y \wedge f(x) = z) \rightarrow y = z)$.

Tal vez el dialetheista pueda asegurar la validez de esta condición en LP entendiendo '$\rightarrow$' como un condicional material pero, al no satisfacer la ley de modus ponens, '$\rightarrow$' no garantizaría la unicidad de los valores que una función le asigna *a una dialetheia*.

En respuesta, Priest (1993, pp. 52-3) decide coger el toro por los cuernos y aceptar que esto es así, aunque minimiza las consecuencias que se siguen de este hecho. Podemos continuar usando definiciones por casos, pero hemos de ser conscientes de que estas definiciones no siempre tendrán un carácter *funcional*, a veces tendrán un carácter meramente relacional, como ocurre con la definición de la evaluación semántica de una oración que, aunque no sea funcional, se define *también por casos*. Por otra parte, es muy importante señalar que *sólo* tenemos motivos para dudar del carácter funcional de una definición por casos en contextos *inconsistentes* (i.e., cuando la definición opera sobre cosas que dan origen a dialetheias, como por ejemplo el conjunto de Russell o la paradoja del mentiroso). Cuando una definición por casos opera sobre elementos consistentes el dialetheismo *no impide* que tenga carácter funcional y respete la condición: $\forall xyz((f(x) = y \wedge f(x) = z) \rightarrow y = z)$. Casi podríamos decir, imitando a Priest, que ciertas definiciones de relaciones son "cuasi-funcionales".

No se puede negar que la respuesta de Priest con respecto a la definición por casos de funciones es bastante coherente con el conjunto de su teoría y no más alarmante o ilegítima de lo que lo pueda ser su tesis sobre la existencia de leyes de inferencia cuasi-válidas. Sin embargo, en este caso tenemos un "pero" añadido. La maniobra de Priest *sí* resta poder *expresivo* a su teoría, ciertas afirmaciones dialetheistas resultan difícilmente expresables una vez introducimos las modificaciones propuestas por Priest para salvar la paradoja reforzada que supone $\underline{\delta}$. Las razones de ello, que expondremos brevemente, se pueden encontrar en Bromand 2002.

Según Bromand, la teoría de Priest puede expresar sin problemas el hecho de que una oración sea verdadera, falsa o verdadera y falsa, pero tropieza con dificultades insalvables a la hora de expresar la idea de que una oración es *sólo* verdadera (o *sólo* falsa). Bromand nos pide que consideremos el siguiente principio dialetheista:

(PD) *Toda oración es sólo verdadera o sólo falsa o verdadera y falsa.*

A continuación construye un argumento según el cual: (a) Si (PD) se pudiese expresar de un modo satisfactorio en LP, entonces podríamos inferir paradojas reforzadas. (b) Sólo podemos evitar paradojas reforzadas sacrificando la posibilidad de expresar (PD) en LP de forma adecuada. Bromand ofrece dos formalizaciones de (PD) en las que v aparece como la *relación* que evalua las oraciones de LP:

(+) $v(p, 1) \wedge \forall x(v(p, x) \rightarrow x = 1) \vee$
$v(p, 0) \wedge \forall x(v(p, x) \rightarrow x = 0) \vee$
$v(p, 1) \wedge v(p, 0) \wedge \forall x(v(p, x) \rightarrow (x = 1 \vee x = 0))$[100]

[100] La otra formalización de (PD) es más simple que (+):

(+') $\forall x(v(p, x) \leftrightarrow x = 1) \vee$
$\forall x(v(p, x) \leftrightarrow x = 0) \vee$
$\forall x(v(p, x) \leftrightarrow (x = 1 \vee x = 0))$

Ambas formulaciones son equivalentes en términos clásicos pero no en la lógica dialetheista, en LP se cumple (+') $\rightarrow$ (+), pero no (+) $\rightarrow$ (+'). Demostrar la segunda implicación requiere la validez de un principio que sólo es *cuasi*-válido para Priest: $(\alpha \wedge \beta) \rightarrow \gamma \vdash \alpha \rightarrow (\beta \rightarrow \gamma)$. Bromand apunta que existe una manera aún más simple de expresar (PD) valiéndonos de la función d que asigna a cada oración, p, el conjunto de sus evaluaciones semánticas: $d(p) = \{x : v(p, x)\}$ (de hecho, esto es lo mismo que hace Priest en 5.2.3.1: definir una función a partir de una relación evaluativa para simplificar la exposición de LP.):

(+'') $d(p) = \{1\} \vee d(p) = \{0\} \vee d(p) = \{0, 1\}$

Bromand se centra en (+) pero advierte que sus observaciones se aplican igualmente a (+') y (+'').

Partiendo del esquema tarskiano [$v(p, 1)$ sii p] y de la siguiente formulación de la paradoja reforzada del mentiroso antes examinada (i.e., 'Esta oración es sólo falsa'):

δ sii $v(\delta, 0) \wedge \forall x(v(\delta, x) \rightarrow x = 0)$

es posible demostrar un absurdo. Sea ($+_\delta$) una ejemplificación de (+) en la que la variable x es sustituida por la oración δ, es posible demostrar entonces que $1 = 0$. Supongamos que se cumple el *segundo* miembro de la disyunción en ($+_\delta$), dado que ese miembro es *idéntico* a δ, tenemos (por el esquema tarskiano) $v(\delta, 1)$ y (por $\forall x(v(\delta, x) \rightarrow x = 0)$ tenemos $1 = 0$. Supongamos que se cumplen el *primer* o *tercer* miembro de la disyunción en ($+_\delta$), entonces δ es verdadero: $v(\delta, 1)$ y (por $\forall x(v(\delta, x) \rightarrow x = 0)$ volvemos a tener $1 = 0$. Como no hay más posibilidades, concluimos que $1 = 0$ y, por tanto, que toda oración es verdadera puesto que todas las posibles evaluaciones de una oración son iguales a 1.[101] Priest intenta evitar este resultado expresando (PD) de otro modo:[102]

(P+) $[(v(p, 1) \wedge \neg v(p, 0)) \vee (v(p, 0) \wedge \neg v(p, 1)) \vee (v(p, 1) \wedge v(p, 0))] \wedge [\forall x(v(p, x) \rightarrow (x = 1 \vee x = 0)]$

[101] Uno podría criticar el argumento de Bromand diciendo que el condicional que aparece en (+), y en el resto de formalizaciones de (PD), es *material* y no satisface por tanto la ley de modus ponens (MP) sin la cual no podemos inferir $1 = 0$ a partir de $v(\delta, 1)$ y de $\forall x(v(\delta, x) \rightarrow x = 0)$. Sin embargo, no cuesta nada sustituir dicho condicional por uno que no sea material y que satisfaga MP (un "entrañamiento", véase 5.2.3.1). Una vez hecho el cambio, volvemos a inferir $1 = 0$. Es importante señalar que cualquier condicional que aparezca en (+) *debe* satisfacer MP, de lo contrario no conseguiría *expresar* un principio *general* como (PD).

[102] Véase Bromand 2002, p. 746.

Sin embargo Bromand responde acertadamente que (P+) no consigue expresar *plenamente* (PD). Según (P+), por ejemplo, un enunciado, p, es sólo verdadero sii

(SV?) $v(p, 1) \wedge \neg v(p, 0) \wedge (\forall x(v(p, x) \rightarrow (x = 1 \vee x = 0))$

Según (SV?), "p es verdadero, no es falso y es verdadero o falso", pero eso *no es lo mismo* que "p es *sólo* verdadero", como expresa el primer elemento de la disyunción (+)

(SV) $v(p, 1) \wedge \forall x(v(p, x) \rightarrow x = 1)$

En lógica clásica (SV) y (SV?) son equivalentes, pero en la lógica paraconsistente de Priest (SV) no se sigue de (SV?). Debido al fallo de la ley de silogismo disyuntivo, no se cumple en general (SV?) $\rightarrow$ (SV). Así pues, (SV?) es demasiado débil para expresar (SV), "p es *sólo* verdadera". Priest se enfrenta a un dilema: O admite la posibilidad de expresar oraciones como (+) en LP y entonces puede demostrar un absurdo; o prohibe la expresión de (+) y prescinde (por razones *ad hoc*) de la forma más plausible y satisfactoria de expresar (PD), uno de los principios básicos de su teoría.[103]

[103] Existe aún otro hecho que nos ayuda a constatar las limitaciones expresivas de LP (la lógica paraconsistente de Priest). En lógica clásica sabemos que $(\alpha \vee \neg\alpha) \leftrightarrow 1$ y $(\alpha \wedge \neg\alpha) \leftrightarrow 0$ ('1' y '0' son aquí constantes proposicionales que representan la verdad y la falsedad respectivamente). Se puede decir que la lógica clásica contiene fórmulas que *expresan* los valores de verdad clásicos (i.e., fórmulas que *siempre* reciben la *misma evaluación*). En LP no parece haber fórmulas α, β, γ tales que $\alpha \leftrightarrow \{1\}$, $\beta \leftrightarrow \{0\}$ y $\gamma \leftrightarrow$

5.3.6 Críticas al Principio de Solución Uniforme (PSU)

Como ya señalamos en 4.2 y en 5.2.1, uno de los argumentos que Priest esgrime a favor del dialetheismo se basa en el Esquema de Inclusión (EI). Todas las paradojas de inclusión comparten la misma estructura, una estructura que revela además los presupuestos y mecanismos básicos que intervienen en la inferencia de una contradicción. EI muestra que, desde cierto nivel de abstracción, el problema que esconden todas las paradojas de inclusión es el *mismo*. Para Priest este hecho es relevante a la hora de abordar la tarea de su solución. Una solución satisfactoria debe resolver todas estas paradojas del *mismo* modo: "same kind of paradox, same kind of solution".[104] Priest llama a esta máxima Principio de Solución Uniforme (PSU) y la justifica diciendo que sólo una solución homogénea a todas las paradojas de inclusión abordará las causas reales del problema sin perderse en aspectos irrelevantes.[105] A juicio de Priest sólo el dialetheismo ofrece una solución de estas características: "aceptar en *todos* los casos la verdad de la

---

$\{0,1\}$ (ni siquiera fórmulas tales que $\alpha \leftrightarrow 1$ $\beta \leftrightarrow 0$). Esto muestra que LP *no puede expresar* sus valores de verdad ya que no existe en LP una fórmula que sea *lógicamente equivalente* a una constante que represente un valor de verdad dialetheista (agradezco este comentario a J. P. Úbeda). Esto guarda relación con otro hecho. A diferencia de lo que ocurre en la lógica clásica (como observábamos en 5.3.1), en LP no existen rasgos sintácticos que determinen cuándo una fórmula α es aceptable ($1 \in v(\alpha)$) o rechazable ($1 \notin v(\alpha)$), i.e., cuándo es equivalente a la verdad y a la no verdad respectivamente. LP sólo puede expresar sintácticamente su noción de absurdo ($\perp$), a saber: $\forall x\ V(\underline{x}) \leftrightarrow \perp$.

104 *BLT*, p. 287, véase también la cita de 4.2.

105 "Any solution that can handle only some members of the family is bound to appear somewhat one-eyed, and as not having got to grips with the fundamental issue." (Priest 1994a, p. 33; y *BLT*, p. 166.)

contradicción que se infiere", así pues, sólo el dialetheismo ofrece una solución satisfactoria a las paradojas.[106]

Además de poco convincente, este argumento no es concluyente. De ser cierto, sólo demostraría que, *actualmente*, no existe ninguna solución consistente "uniforme" a las paradojas de inclusión. Pero no se sigue de ahí que sea *imposible* construir tal solución o que haya algo relacionado con la consistencia que nos impida plantearnos el proyecto de ofrecer una solución consistente uniforme. Varios lógicos consistentes, empezando por el propio Russell, han propuesto soluciones uniformes, obviamente buscamos una que sea satisfactoria y la de Russell no lo es, pero ¿acaso lo es la de Priest? Sin embargo, los problemas fundamentales del argumento de Priest a favor de PSU son otros. En su justificación de PSU, Priest recurre a una analogía poco convincente. Según Priest, si dos personas tienen el *mismo tipo* de enfermedad, serán curadas del *mismo* modo. Tal vez reciban tratamientos diferentes (distintos medicamentos, etc.), pero en la medida en que ambas curas ataquen la *causa común* de la enfermedad, ambas serán del *mismo tipo* o *clase*.[107] Como justificación esta analogía resulta pobre y poco iluminadora.

El ejemplo que ofrece Priest (en la nota anterior) es pobre entre otras cosas porque a veces habla de enfermedades del *mismo tipo* y otras

---

[106] Véase en este sentido la cita de *BLT* en la nota 9, esta afirmación aparece ya en Priest 1994a, p. 33.

[107] "[S]ame kind of illness, same kind of cure –if two people have the same illness, they are to be cured in the same way. In a superficial sense, this is obviously false. For example, the same illness can be treated with two different drugs. But in a more profound, and more important, sense, the principle is clearly correct: if we have one illness in the two people, this must be due to the same cause. So the two people must be cured in the same way, namely, by attacking that cause." *BLT*, pp. 287-8.

de la *misma* enfermedad. Las divergencias "taxonómicas" que pueden existir entre lo que tienen dos personas con la *misma* enfermedad (por ejemplo, la gripe) son mínimas en comparación con las divergencias que podemos imaginar entre paradojas tan distintas como la del mentiroso y la de Russell. En los capítulos 2 y 3 vimos que existen subcategorías importantes dentro de las paradojas de inclusión. En cierto modo sólo "la paradoja del mentiroso" es ya una colección de paradojas distintas aunque "del mismo tipo" –existen diferencias significativas en relación con los medios empleados para conseguir la autorreferencia (nombres, gödelización, etc.), el número de oraciones involucradas (cadenas de mentirosos), o de principios semánticos, etc.–. El ejemplo de Priest sería más justo si en vez de hablar de la *misma* enfermedad hablase sólo de enfermedades del *mismo tipo*. En ese caso podríamos preguntar *de qué tipo* o *en qué aspecto se parecen*, al igual que nos preguntamos en el caso de las paradojas relevantes. Esta pregunta es fundamental porque determina el contenido y grado de abstracción de nuestro análisis.

Clasificar es una tarea consistente en agrupar objetos a partir de las similitudes que podemos encontrar entre ellos. Pensar que un objeto *a* es "de clase C" supone abstraer aquellas propiedades que comparte *a* con todos los objetos de clase C y dejar de lado el resto. Toda propiedad, P, que imaginemos (por artificial, vacua o rebuscada que sea) nos permite concebir o caracterizar una *clase* (paradojas aparte), "la de todos los objetos que satisfacen P". En este sentido, definir "clases de objetos" es en buena medida una tarea arbitraria. Es cierto que, una vez definidas o identificadas, decidir si tal objeto pertenece a tal clase no es un ejercicio arbitrario, pero decidir en general qué aspectos de un objeto vamos a utilizar para clasificarlo sí requiere nuestro arbitrio. Siempre encontraremos una clase que agrupe exactamente a los elementos que nos interesa agrupar y siempre encontraremos una clase en la que encuadrar un objeto particular. La predicación es, sin ir más lejos, una forma de clasificar objetos y, del mismo modo que no podríamos responder a la pregunta "¿Cuántas cosas podemos predicar de algo?", tampoco podríamos responder a preguntas como "¿Cuántas clases de objetos hay?" o "¿A cuántas clasificaciones pertenece este objeto?". Hay clases y clasificaciones para todos los gustos dependiendo, en parte, de nuestros intereses y necesidades. Así pues, cuando alguien nos dice que dos cosas (dos enfermedades o dos paradojas) son del mismo tipo o clase, cabe preguntar "*con respecto a qué*" puesto que desde cierta perspectiva dos cosas cualesquiera pueden ser "del mismo tipo o clase" y, desde otra, "de distinto". Desde el punto de vista de los sistemas y órganos afectados, la gripe y la tuberculosis son enfermedades *respiratorias* y el sida, en cambio, no; pero desde el punto de vista de la naturaleza del organismo que las produce, la gripe y el sida son enfermedades *víricas* y la tuberculosis *bacteriana*.

Como Priest está interesado fundamentalmente en la *solución de problemas*, la clasificación que sugiere, tanto para las paradojas como para las enfermedades, agrupa los objetos relevantes identificando en ellos elementos similares que *causen* problemas similares. No nos interesan clasificaciones del tipo "enfermedades surgidas en el siglo XX", sino clasificaciones del tipo "enfermedades *víricas*". Al clasificar una enfermedad como vírica identificamos la *causa* del problema (el tipo de organismo que origina la enfermedad) y sugerimos también la mayor o menor eficacia de ciertos tratamientos terapéuticos y medidas preventivas. En el caso de las paradojas que nos preocupan y del problema que suponen, más que a *causas naturales* que lo justifiquen, debemos apelar a *razones conceptuales* (*BLT*, p. 288). La clasificación relevante propuesta por Priest es la de "paradojas de inclusión", una clasificación que postula la existencia de un *patrón estructural* común a todas ellas que, por su constitución, origina la inferencia de contradicciones. Esta clasificación revela los elementos fundamentales que articulan el problema (Ω y δ) y, sobre todo, las relaciones estructurales que mantienen y que justifican la inferencia de una contradicción (δ actúa sobre Ω generando un elemento que está y no está en Ω). En este sentido Priest afirma que EI *explica* el *problema* que suponen las *paradojas* y el dialetheismo le da una solución *uniforme*.

Llegados a este punto tropezamos con otra dificultad del ejemplo propuesto por Priest.[108] Al igual que toda clasificación, todo análisis estructural de situaciones similares y toda formulación de un problema (y de su solución) suponen un ejercicio de abstracción que puede tener varios niveles de profundidad. Un análisis estructural muestra que 'Juan bebe vino' y 'Juan bebe agua' tienen una estructura idéntica: Beber(x, y), diferente de la estructura de 'Juan es hermano de Pedro': Hermano(x, y). Un análisis estructural más abstracto muestra, en cambio, que las tres oraciones tienen la misma estructura: R(x, y). Otro tanto ocurre con las

[108] Esta objeción (o una muy similar) se puede encontrar en Smith (2000).

clasificaciones. Los chimpancés no son humanos, sin embargo, ambos son primates a diferencia de los elefantes que, sin embargo, son mamíferos, al igual que chimpancés y humanos. Lo mismo se puede decir de la formulación de un problema y de su solución. Un *mismo* problema abstracto puede resolverse resolviendo *distintos* problemas de carácter más concreto. Podemos resolver "el problema de la despoblación de un territorio" resolviendo "el problema del descenso de la natalidad" o resolviendo "el problema de los flujos migratorios".[109]

Algunos análisis estructurales y categoriales nos permiten ascender y descender en "orden de abstracción". Ofrecer una solución de la misma clase a problemas de la misma clase o con la misma estructura exige que el nivel de abstracción de la solución y del problema analizados sea aproximadamente el *mismo*. La solución *concreta* que ofrecemos a la paradoja del mentiroso involucra principios y conceptos semánticos y no es por tanto igual a la solución *concreta* que ofrecemos a la paradoja de Russell, que no involucra *ese tipo* de conceptos. Para decir que ambas soluciones son del mismo tipo debemos situarnos en un nivel de abstracción adecuado y debemos apelar a una formulación del problema que permita ver la uniformidad de fondo de soluciones aparentemente distintas.

Una vez situados al nivel de abstracción que marca el Esquema de Inclusión, nada hay en principio que impida al lógico clásico decir que,

[109] No existe una única manera de clasificar jerárquicamente problemas, soluciones, categorías o estructuras por "orden de abstracción" y tampoco podemos asignar un "grado de abstracción" a una categoría o a la descripción de una estructura, problema,

en un sentido abstracto y general, todas sus soluciones a las distintas paradojas son soluciones *uniformes* a un problema *del mismo tipo*. Lo importante en este caso es identificar *cuál* es el problema y ofrecer una solución al problema descrito. Para el lógico consistente el *problema* de las paradojas de inclusión es la inferencia de una contradicción a partir de la relación que mantienen entre sí los elementos de EI y una solución *uniforme* del mismo pasa por rechazar la contradicción bloqueando por *reductio* alguno de los elementos de EI. *Eso* es lo que él hace con *todas* las paradojas de inclusión, aunque apele a diferentes recursos o bloquee diferentes elementos del esquema atendiendo a las particularidades de cada caso (i.e., a lo que el esquema *no capta*). Alguien podría quejarse diciendo que debería bloquear siempre el *mismo* elemento, pero esta acusación no tiene base. De lo que se trata es de *identificar correctamente* el problema y de darle una solución *general.*

El problema de las enfermedades víricas no es necesariamente la presencia de un virus en nuestro organismo sino el mecanismo por el cual el virus infecta a las células reproduciéndose y destruyéndolas sin que podamos evitarlo. Atacar del mismo modo la *causa* de una enfermedad vírica no tiene por qué consistir pues en *matar el virus*. Alguien que consigue matar un virus con antibióticos, alguien que genera defensas en un organismo mediante vacunas (u otros medios) y alguien que inhibe la reproducción del virus mediante medicamentos están ofreciendo soluciones "*del mismo tipo*" a un problema "*del mismo tipo*". Todos

---

etc. con precisión matemática. Estas cuestiones son relativas, dependen del contexto y son siempre vagas y bastante arbitrarias.

ofrecen "soluciones que atacan la *causa* del problema (i.e., la reproducción letal de un virus en nuestro cuerpo)". No ofrecen soluciones que, como rezar o ignorar el dolor, no atacan la causa del problema.

Priest dice que ofrece la *misma* solución en *todos* los casos: *aceptar las contradicciones*, pero también esta afirmación es relativa. Aceptar contradicciones en el caso de la paradoja de Russell requiere tomar decisiones distintas a las que adoptamos para aceptar las contradicciones que impone la paradoja del mentiroso. Desde cierto nivel de abstracción, las soluciones dialetheistas a la paradoja del mentiroso y a la paradoja de Russell también son *distintas*. En resumen, el problema del argumento de Priest a favor del dialetheismo es que aunque PSU fuese un principio correcto, no ofrecería argumentos decisivos para decantar la balanza: O PSU es un principio trivial y entonces tanto el lógico consistente como el dialetheista lo cumplen (aunque divergen a la hora de identificar y describir el problema de fondo); o es un principio muy restringido y entonces su validez es más que discutible.

Sin embargo, el problema más importante de PSU es a nuestro juicio otro. Una cosa es afirmar que EI encierra las condiciones necesarias y suficientes para obtener una contradicción de inclusión y otra muy distinta afirmar que EI encierra las condiciones necesarias y suficientes para obtener una *paradoja* de inclusión. Grattan-Guinness criticaba el Esquema de Inclusión por no ofrecer condiciones necesarias y suficientes para identificar a las paradojas reflexivas (véase 2.8.2). Juzgamos entonces que EI *sí* ofrecía condiciones necesarias para identificar paradojas de inclusión, pero también consideramos que EI *no* ofrecía condiciones *suficientes* y que no tenía por qué ofrecerlas. Las razones que ofrecíamos allí apelaban al hecho de que EI refleja sencillamente *la estructura formal de un tipo de razonamientos: las "contradicciones de inclusión"*, pero el hecho de que una contradicción de inclusión se perciba como paradójica no depende tanto de su estructura como de nuestra disposición a aceptar la contradicción que nos impone. Éste era el mayor punto de desacuerdo con respecto a las tesis de Priest sobre EI en *BLT*.

En el caso de la "paradoja" del barbero, nuestra disposición a aceptar la contradicción es nula, en el caso del teorema de Gödel los lógicos consistentes rechazan la contradicción y los dialetheistas no. Vimos en 2.7 que el teorema de Cantor también encierra una contradicción de inclusión en su demostración, sin embargo, en este caso ni Priest ni el lógico clásico aceptan la verdad de la contradicción. Lo relevante de esta discusión es percatarse de que *no hay rasgos estructurales o formales* que justifiquen el hecho de que un razonamiento sea *paradójico*. Lo que hace que el razonamiento sea paradójico es nuestra percepción de que sus premisas son verdaderas, su conclusión (sólo) falsa y las reglas de inferencia aplicadas, válidas. Estas observaciones subrayan curiosamente lo defendido por el propio Priest en 5.3.1: No existe una descripción estructural que ofrezca condiciones necesarias y suficientes de *lo paradójico* o de *lo rechazable*. La "paradoja" del barbero y el teorema de Cantor encierran contradicciones de inclusión, pero ninguna de ellas es una paradoja (o, al menos, no para Priest), ambas contradicciones se resuelven mediante una reducción al absurdo. Las implicaciones de este hecho son extraordinariamente relevantes para la presente discusión. Aunque PSU fuese un principio válido, el dialetheista no podría apelar a él (pero sí, curiosamente, el lógico clásico). El dialetheista reconoce que *algunas* contradicciones de inclusión *son* paradójicas y otras, pese a compartir la *misma* estructura, *no*. Presumiblemente sólo las contradicciones de inclusión paradójicas son verdaderas, el resto son sólo falsas y rechazables. Desde la perspectiva del lógico clásico *todas* las contradicciones de inclusión son rechazables mediante una *reductio* de alguno de sus principios, lo único que diferencia a las paradojas de inclusión de las meras contradicciones de inclusión es la facilidad que tenemos a la hora de identificar el elemento sobre el que opera la *reductio*. Así pues, el lógico clásico ofrece una solución *uniforme* (una *reductio*) al problema que encierran *todas* las contradicciones de inclusión (i.e., la inferencia de una contradicción), mientras que el lógico dialetheista ofrece dos soluciones: *algunas* contradicciones de inclusión son verdaderas (las paradojas) y *otras*, en cambio, son sólo falsas y rechazables mediante una *reductio*.

5.3.7 El Teorema de Gödel no muestra la inconsistencia de nuestro razonamiento

¿Apoya el teorema de Gödel (TG) la inconsistencia de nuestra teoría intuitiva de la demostración en aritmética y en otros sistemas formales equiparables? Chihara (1984b) ha criticado las conclusiones de Priest basándose sobre todo en dos objeciones. La primera es de carácter escéptico, pone en duda que Priest haya ofrecido una formalización correcta de "nuestra teoría intuitiva de la demostración". La segunda es antimecanicista, ataca la idea (defendida en el argumento (2) de 5.2.1.1) de que nuestra noción *intuitiva* de demostración sea recursiva y, por tanto mecanizable. Si Chihara tiene razón, entonces la "paradoja de Gödel" se resuelve del modo tradicional: Intuitivamente podemos inferir verdades (la oración *g* en 5.2.1.1) indemostrables en sistemas axiomáticos, *ergo* nuestra teoría intuitiva de la inferencia *no* es formalizable, axiomatizable o mecanizable.

(1) *Objeción escéptica*: Priest afirma que los axiomas y reglas que rigen nuestra noción intuitiva de demostración (así como la propia noción) pueden ser representados mediante una teoría formal T intuitivamente correcta, aunque inconsistente. Chihara (op. cit., p. 119) se pregunta si el argumento de Priest requiere (a) que T sea *realmente* correcta o (b) que sólo lo sea *aparentemente*. (b) es insuficiente, ya que T podría ser incorrecta y la "teoría correcta", T', podría ser consistente. (También Priest utiliza en su rechazo de PNC argumentos que apelan a la falibilidad histórica de ciertos principios y teorías. Priest 2003 se centra en esta cuestión) Por otra parte, si Priest necesita (a), entonces debe establecer dos afirmaciones. *Primera*: "Si T formaliza correctamente nuestra teoría de la demostración, T sólo demuestra verdades". *Segunda*: "Podemos demostrar que T formaliza correctamente nuestra teoría de la demostración". Según Chihara (ibíd, pp. 120-1), la primera afirmación es correcta, pero Priest no demuestra la segunda. Así pues, no podemos determinar la corrección de T, lo que nos enfrenta a la posibilidad de que sólo sea aparentemente correcta y a las consecuencias que de ello se

siguen. Aunque T sea correcta y sólo demuestre verdades, ello no significa que podamos demostrar (o siquiera saber) que lo es, tal y como exige la segunda afirmación, por tanto, no sabemos si (a) es el caso o si lo es (b).

(2) *Objeción antimecanicista*. Priest defiende que nuestros axiomas y mecanismos intuitivos de inferencia se pueden formalizar y describir en T de forma recursiva, lo que justifica que podamos aplicar TG sobre T, Chihara (íbid. 121-3) señala que esta tesis no es obvia en términos clásicos o intuicionistas. Gödel y los "antimecanicistas", por ejemplo, la rechazaron. Según ellos, TG muestra que no puede existir una axiomatización completa del razonamiento humano que permita describir de modo recursivo la totalidad de nuestras prácticas de demostración. No existe un procedimiento mecánico que nos permita decidir qué es un axioma y qué no lo es. Chihara también duda de la posibilidad de acomodar la aritmética infinitaria y PEV (el esquema tarskiano) en T (íbid., p. 123). Por todo ello, Chihara defiende el principio de no contradicción (PNC) y rechaza la tesis mecanicista con respecto a nuestro razonamiento intuitivo.

La objeción escéptica de Chihara (1) es poderosa, pero no sólo juega en contra del dialetheista sino también del matemático clásico. Priest no ha demostrado que su teoría inconsistente de la demostración, T, sea la correcta, pero, para refutar T, el matemático clásico *también* ha de demostrar que su teoría es la correcta, y lo ha de hacer sin apelar a PNC. En caso contrario, la teoría consistente podría ser sólo aparentemente correcta y la teoría correcta podría revelarse inconsistente. Así pues, (1) deja la partida en tablas.

La objeción antimecanicista (2) sí pretende inclinar la balanza. Pero, para ser efectiva, no puede incurrir en una petición de principio, así pues, no puede limitarse a la tesis (AM): "TG demuestra que nuestra noción intuitiva de demostración no es recursiva (y, por tanto, que nuestros mecanismos intuitivos de inferencia no son mecanizables), *porque, si lo fueran, demostraríamos contradicciones*". AM rechaza la posible mecanización del razonamiento apelando exclusivamente a la validez de PNC, pero no ofrece argumentos a favor de PNC (*IC*, p. 53).[110] Chihara también apela a Gödel para rechazar la tesis mecanicista, la idea parece ser la siguiente: (I) Nuestras teorías intuitivas de la demostración, T, son revisables, podemos añadir nuevos axiomas para inferir (nuevas) verdades contrastadas. Por ello ninguna descripción de nuestros mecanismos inferenciales puede ser completa. (II) *En particular*, TG *nos muestra en qué sentido toda teoría* T *es incompleta y revisable*. Si T es consistente (y satisface los requisitos de TG), tropezamos con una oración de Gödel, *g* (véase 5.2.1.1), verdadera pero indemostrable. Eso nos fuerza a añadir *g* a T en calidad de *axioma*, generando, así, una nueva teoría T'. Ahora bien, si T' es consistente (etc.), por TG, existirá una oración *g*' verdadera y no demostrable que deberá añadirse a T' para

[110] En 5.2.1.1, Priest da razones a favor de la recursividad de nuestra noción de demostración. A éstas, Priest añade ahora que tanto los lógicos clásicos como los intuicionistas están de acuerdo en la idea de que debemos poder *decidir* qué es una prueba y qué no lo es, lo que sugiere la posibilidad de un test mecánico (*IC*, pp. 52-3). (Aunque no es menos cierto que ambos están en desacuerdo con respecto a *qué* pruebas

obtener una nueva teoría T''. El resultado de repetir esta operación es una jerarquía de teorías en la que ninguna teoría T* consigue axiomatizar por completo los razonamientos intuitivos que nos permiten inferir la verdad de la correspondiente oración de Gödel g^*.

Esta objeción consta de una observación general, (I), y de un argumento particular, (II), que ilustra la verdad de (I). Podemos atacar (II) por dos motivos. En primer lugar, porque (II) presupone la interpretación de TG que debe justificar. Al añadir g como axioma a T, para obtener T' *presupone* que *g no es demostrable en* T. Ahora bien, la *única* razón que alega para ello es la *consistencia* de T, que es precisamente lo que debe justificar frente al dialetheista (en resumen, (II) incurre en una petición de principio). En segundo lugar, (II) es dudoso porque el argumento que ofrece para mostrar que los axiomas de cualquier T_α en la jerarquía son incompletos es siempre *el mismo* (*IC*, p. 54). (II) ofrece una estrategia *mecánica* o *recursiva* para "mostrar" la verdad de cualquier g de la jerarquía. Esto nos hace sospechar de la "novedad" real de g y de su carácter axiomático. Si buscamos una teoría suficientemente general, g parece más bien una oración demostrable. Por otra parte, vimos en 4.2, 4.3.2 y 4.5.1 que las teorías jerárquicas no están exentas de problemas. Es difícil teorizar coherentemente sobre *todas* ellas.

Las objeciones de Priest a AM y (II) son convincentes. Sin embargo, no bastan para excluir la verdad de (I), que, en definitiva, viene a poner sobre la mesa el argumento escéptico de Chihara. El principal

son aceptables, lo que implica que no existe consenso con respecto a *cuál* es la teoría correcta de la demostración.)

problema del uso que hace Priest de TG es que lo aplica a una teoría, T, que supuestamente formaliza de forma correcta nuestras técnicas intuitivas de demostración y, como señala Chihara, no está clara ni la existencia de T ni su alcance (ni, por tanto, que sea axiomatizable) dada la evolución histórica de nuestras ideas sobre la demostración.[111] Nuestras teorías intuitivas son falibles y revisables, T podría resultar incompleta y, si eso es así, nada nos garantiza que el resultado de completar sus axiomas dé como resultado una teoría inconsistente. Tal vez "la teoría correcta" sea consistente, tal vez no. A menos que resolvamos esta cuestión TG será neutral.[112]

5.4 Conclusión: ¿Es el dialetheismo una solución satisfactoria a las paradojas?

Creo que, una vez examinados los argumentos de 5.3, la respuesta es "No". El dialetheismo es una teoría que tropieza con serios problemas de coherencia a la hora de exponer, con la generalidad deseada (y prometida), sus tesis y principios fundamentales. En el caso del lógico consistente estos problemas derivan de su incapacidad para ofrecer una

[111] Priest (*IC*, p. 54-5) insiste en que él examina la cuestión sincrónicamente, pero dice que diacrónicamente también podrían seguirse los mismos resultados. Sin embargo, no está en absoluto claro que esto sea posible.

[112] Recientemente, S. Shapiro (2002) ha mostrado que es posible añadir la semántica dialetheista de Priest a la aritmética de Peano, PA, para producir un sistema formal PA* que: (1) es axiomatizable de forma *recursiva*; (2) contiene su propio predicado de verdad; (3) es inconsistente (pero no trivial), ya que contiene una prueba de *g*, la oración de Gödel; (4) puede demostrar su corrección [*soundness*]. Sin embargo, PA* tiene consecuencias no deseables (incluso para un dialetheista, según Shapiro): (1') existen contradicciones en el más elemental lenguaje de la aritmética (PA y la aritmética de Robinson son inconsistentes); (2') se puede demostrar que existe un número *n* que es y no es el código de la demostración de *g*, la oración de Gödel.

teoría *general consistente*; en el caso del dialetheista estos problemas derivan de su incapacidad para ofrecer criterios claros, una vez rechazado PNC, para decidir cuándo una proposición es rechazable o absurda y cuándo no, lo que no permite excluir con seguridad la posibilidad del *trivialismo*. Muchas de las tesis que Priest acepta como verdaderas parecen tener consecuencias dialetheistas que van, no sólo a favor de su falsedad (en cuyo caso serían dialetheias), sino *en contra de su verdad*. Para evitar problemas de este tipo, Priest toma a menudo decisiones *ad hoc* que no difieren mucho de las que toma el lógico consistente y que generan problemas similares. Entre estas decisiones encontramos la distinción entre falsedad y no-verdad; la aceptación de las dialetheias, eludiendo la obligación pareja de rechazarlas; o el reconocimiento de (la verdad de) tesis que niegan (la verdad de) principios dialetheistas cruciales (por ejemplo, "No hay oraciones verdaderas y no verdaderas"). Por otra parte, muchas de las decisiones *ad hoc* que toma el dialetheista nos fuerzan a introducir restricciones inferenciales (ECQ, SD, MP, ...) y expresivas (algunas funciones definidas por casos y paradojas reforzadas como 'Esta oración es *sólo* falsa') muy similares a las que nos imponen las soluciones consistentes conocidas. Si unimos a estas observaciones el hecho de que el dialetheismo *no muestra la imposiblilidad* de resolver el problema de las paradojas de forma consistente (ni los argumentos del capítulo 4, ni los relativos al teorema de Gödel establecen tal resultado), nos veremos forzados a admitir que el dialetheismo no zanja el problema que suponen las paradojas y tampoco las elimina.

Presumiblemente, ofrecer una solución satisfactoria a un problema consiste en proporcionar una solución que cancele la búsqueda de *otra solución mejor*. Desde esta perspectiva, pocos serán seguramente los problemas relevantes que se hayan resuelto de forma satisfactoria. Sin embargo, la satisfacción es también una cuestión de grados. Algunos problemas quizá nunca lleguen a tener una solución plenamente satisfactoria, pero sí una solución aceptablemente satisfactoria. Tal vez las paradojas que nos ocupan ni siquiera lleguen a disfrutar de lo segundo. Todas las teorías que buscan soluciones consistentes a las paradojas han tropezado históricamente con la persistencia de paradojas reforzadas y, por tanto, con inconsistencias. Acabamos de ver que el enfoque inconsistente más sofisticado y meritorio del problema también contempla paradojas reforzadas (paradojas que cuestionan la verdad de los principios dialetheistas). Así pues, no hemos resuelto el problema satisfactoriamente y tampoco lo hemos resuelto *más* satisfactoriamente que el lógico consistente. Como dijimos al principio de este trabajo, una paradoja es algo que nos sentimos inclinados a rechazar pese a desprenderse necesariamente de convicciones y principios de inferencia de cuya verdad nos sentimos seguros. Esta experiencia es común al lógico consistente y al dialetheista, aceptar algunas contradicciones no disuelve el problema. En la medida en que rechacemos el trivialismo, algunas contradicciones no podrán ser verdaderas, y hemos visto a lo largo de 5.3 que tenemos sobrados motivos para considerar que el dialetheismo implica contradicciones inaceptables incluso para el dialetheista.

¿Es esto una demostración de que *no existen contradicciones verdaderas*? Por desgracia, *no*. Lo único que hemos hecho en este capítulo es observar que una de las teorías que defienden esta postura (el dialetheismo, la más sólida) es insatisfactoria. Del mismo modo que las críticas de Priest a las teorías consistentes actuales no excluyen la posibilidad de una teoría consistente adecuada, tampoco nuestras críticas excluyen la posibilidad de una teoría inconsistente apropiada. Cualquier teoría inconsistente se enfrentará previsiblemente a problemas parecidos a los aquí comentados y, ciertamente, esos problemas son serios. Sin embargo, no hay que olvidar que el apoyo de Priest al dialetheismo no es caprichoso, su argumento central es tan poderoso como el que aquí hemos ofrecido en su contra, es más, la conclusión es la *misma*: Las teorías supuestamente consistentes desembocan en el trivialismo cuando tienen un alcance general. Cuando estas teorías abordan el problema de las paradojas tropiezan con versiones reforzadas de las mismas, eso significa que no pueden garantizar la *consistencia* global de su solución a las paradojas y significa también que, cuando intentamos extenderlas para que den cuenta de cualquier teoría (*incluidas ellas mismas*) se vuelven inconsistentes. En la medida en que las teorías tratadas respetan ECQ, cualquier inconsistencia desembocará pues en el trivialismo. Siempre podemos restringir las soluciones consistentes a un ámbito determinado, pero esa maniobra siempre estará al alcance de cualquier solución paraconsistente.

Sólo encontrando una solución consistente o paraconsistente satisfactoria podríamos zanjar la cuestión con respecto a la existencia de contradicciones verdaderas. Por desgracia, ninguna teoría satisfactoria en este sentido se vislumbra en el horizonte y, del mismo modo que no podemos descartar la posibilidad de encontrarla en el futuro, tampoco podemos descartar la posibilidad de no encontrarla nunca o, peor aún, de que no exista. Esto plantea un problema que Priest insinúa en algunos textos.[113] ¿Podemos excluir la verdad del *trivialismo*? Si el trivialismo fuese cierto, entonces el dialetheismo resultaría ser una teoría correcta después de todo, aunque eso sí, *también* las teorías consistentes atacadas por Priest serían correctas. El dialetheismo sería una teoría correcta porque *todo* sería verdadero y *toda* teoría correcta. Tal vez éste sea el mayor reto que las paradojas ponen sobre la mesa, un reto muy similar (aunque mucho más radical) al que supone el escepticismo global. Al igual que ocurre con el argumento escéptico podemos ignorar sus resultados en nuestra vida cotidiana porque dependen de la realidad de posibilidades que psicológicamente percibimos como altamente improbables, pero posibilidades que, psicológicamente, tampoco podemos descartar a priori y que resultan relevantes cuando nos planteamos cuestiones generales en epistemología o cuando intentamos explicar qué es el conocimiento. Con las paradojas ocurre lo mismo, afectan a enunciados que podemos ignorar habitualmente, pero que debemos tener en cuenta cuando nos planteamos seriamente la definición de ciertos conceptos básicos (verdad, conjunto ...). Es entonces cuando las paradojas amenazan con hundirnos en el trivialismo o con dejarnos "mudos" del mismo modo en que "enmudeció" Wittgenstein al final del

[113] Sobre todo en Priest 2000b.

Tractatus. Quizá tomarse en serio el problema de las paradojas sea tomarse en serio el problema del trivialismo.

6. Conclusiones

6.1 Recapitulación: El valor real del Esquema de Inclusión.

Russell (1908) denominó paradojas de autorreferencia o reflexividad a un grupo significativo de antinomias que parecían compartir varios rasgos distintivos. El más importante de ellos era la presencia de cierta circularidad viciosa en la definición de alguno de los elementos cruciales en la inferencia de una contradicción. Recientemente, Priest (1994a) ha propuesto una descripción estructural más exacta de estas paradojas basándose a su vez en el análisis estructural unificado de cuatro de ellas (Russell [2], Burali-Forti [7], Mirimanoff [9] y Cantor [10]) que efectuó el propio Russell en 1905.

En el presente trabajo (capítulos 2 y 3) hemos defendido que Russell y Priest tienen razón en lo fundamental: Las paradojas reflexivas tienen una estructura común que puede ser descrita mediante el Esquema de Inclusión (EI) de Russell-Priest. Todas ellas involucran una totalidad absoluta Ω cuyos miembros satisfacen cierta propiedad φ. Ω contiene forzosamente un elemento, $\delta(\Omega)$, que sólo podemos definir circularmente a partir de Ω mediante una operación (un diagonalizador), δ, que genera una contradicción al aplicarse sobre Ω: $\delta(\Omega) \in \Omega$ y $\delta(\Omega) \notin \Omega$. Pero, ¿qué importancia tienen en general las paradojas y el esquema?

Más allá de su apariencia de simples acertijos o pasatiempos intelectuales, las paradojas reflexivas aquí descritas han supuesto hasta la fecha obstáculos insalvables a la hora de construir teorías coherentes de carácter general sobre algunos de nuestros conceptos más básicos: "verdad", "conjunto", "infinito", "satisfacción de una propiedad", "demostración", etcétera. Este tipo de conceptos aparece con frecuencia en las ciencias formales actuales y también en un buen número de ciencias naturales y sociales. No podemos resignarnos a admitir que toda explicación teórica completa de nociones tan básicas haya de ser forzosamente contradictoria, así pues, sin resolver las paradojas no obtendremos una comprensión satisfactoria de esos términos. Desde esta perspectiva, las paradojas reflexivas constituyen un serio problema. Ahora bien, ¿constituyen realmente *un* problema general o constituyen por contra *un conjunto* de problemas *dispares* e independientes? El Esquema de Inclusión ofrece indicios a favor de la primera tesis, ésa es su relevancia. Si alcanzamos un nivel de abstracción apropiado, las paradojas reflexivas se pueden describir como ejemplificaciones de un *mismo problema general.* EI muestra que el razonamiento paradójico tiene *la misma forma* en todos los casos, involucra entidades y operaciones muy similares y se produce siempre del mismo modo. Este hecho no es en absoluto despreciable: en primer lugar, describe las condiciones y circunstancias en las que probablemente tropezaremos con una paradoja; en segundo, nos muestra cuáles son los ingredientes relevantes de la paradoja y cómo se articulan entre sí invariablemente para dar lugar a una contradicción. EI nos permite de hecho describir nuevas paradojas, hablar de forma general de un sinfín de ellas, ver en qué sentido se relacionan unas con otras y alcanzar conclusiones generales. El esquema constituye un paso adelante en nuestra comprensión de un fenómeno complejo, antes sólo teníamos intuiciones más o menos vagas con respecto a la unidad de varias paradojas (de ahí el nombre genérico de "reflexivas"), ahora tenemos una descripción que aporta criterios estructurales más precisos, criterios que permiten concretar nuestras intuiciones de partida identificando los elementos básicos de una paradoja reflexiva y los mecanismos que la desencadenan.

Una de las virtudes del trabajo de Priest (y, con anterioridad, del de Russell) es que logra identificar un objeto abstracto, algo que proporciona unidad interna a una serie de fenómenos dispares y que podemos convertir ahora en objeto de estudio. Ni este trabajo ni cualquier otra investigación destinada a hablar de paradojas reflexivas *en general* habría sido muy precisa (o incluso posible) de no haber existido una manera uniforme y rigurosa de describirlas. Muchos han hablado de paradojas reflexivas antes. Al hacerlo, han hablado en realidad (como Russell en 1908) de la *circularidad viciosa* que acompaña a todas estas paradojas. Pero esta *circularidad* no es un rasgo estructural exclusivo de las paradojas reflexivas. Muchos razonamientos y definiciones manifiestan este tipo de problemas sin llegar a establecer contradicciones. Así pues, EI ofrece una descripción estructural más exacta y, por tanto, notables ventajas teóricas en el estudio de las paradojas y el establecimiento de tesis generales. Pero aún hay más.

En el capítulo 1 (y también en 3.2.13) defendimos que las paradojas reflexivas tienen un interés añadido para la filosofía (Priest 2002 ofrece una excelente defensa de esta idea). Al reflexionar sobre estas paradojas generamos problemas filosóficos relacionados con nuestra comprensión del mundo y de nuestros límites cognitivos. Muchas de las paradojas que acosan a la filosofía son, en cierto modo, "de segundo orden", surgen a partir de reflexiones sobre paradojas más básicas que también abordan las ciencias particulares. Curiosamente, la forma de estas paradojas propias de la filosofía y de la metafísica recuerda a menudo la forma de las contradicciones de inclusión, son reflexiones que se dirigen a totalidades absolutas de las que forman parte necesariamente, pero que pretenden trascender de algún modo en sus conclusiones: Pensamientos que se dirigen hacia lo pensable para imaginar lo impensable; o investigaciones que pertenecen al ámbito de la justificación racional, de lo demostrable o de lo cognoscible y que concluyen, sin embargo, que no es posible justificar racionalmente nuestras creencias, o demostrar, o conocer su verdad. Así pues, el Esquema de Inclusión también describe "grosso modo" la *forma* que tienen estos problemas de orden superior y, en ese sentido, nos ayuda a ver el parentesco que guardan entre sí y con otras paradojas. ¿Por qué no hablar pues de contradicciones de inclusión? Después de todo las descripciones estructurales de carácter general son útiles: Hablamos de argumentos trascendentales para referirnos a ciertas estrategias de argumentación no empíricas, hablamos de círculos viciosos, o de regresos al infinito cuando hablamos de la estructura de ciertas definiciones o argumentos, etcétera.

Pero, pese a su utilidad innegable, EI no ofrece una respuesta a todos los problemas que nos preocupan en relación con las paradojas. Tanto Priest como Russell defienden tesis de las que aquí nos hemos desmarcado. Ambos filósofos piensan (o asumen implícitamente) que los análisis estructurales descritos en el capítulo 2, el Esquema de Inclusión (EI, 2.6) en el caso de Priest y la estructura proscrita por el Principio de Circularidad Viciosa (PCV, 2.2) en el de Russell, aportan condiciones necesarias *y suficientes* para determinar cuándo nos hallamos en presencia de una paradoja reflexiva y cuándo no. Así mismo, ambos autores piensan que la existencia de un paralelismo estructural entre todas estas paradojas justifica también el diseño de una estrategia de solución *única*. Nosotros no suscribimos aquí ninguna de estas tesis (sobre todo la primera, la segunda admite matices) y tampoco nos comprometemos con la existencia de una solución satisfactoria al problema planteado en el primer capítulo.

En ningún momento hemos puesto en duda que Russell y Priest hayan identificado condiciones *necesarias* para la obtención de una paradoja reflexiva. Todas las paradojas que nos preocupan encajan en EI (como afirma Priest) y, por tanto, muestran problemas de circularidad viciosa (como afirma Russell). Sin embargo, *ni* los criterios aportados por Russell *ni* los criterios aportados por Priest *bastan* para obtener una paradoja: *No todo* lo que encaja en EI es una paradoja. EI muestra que las condiciones identificadas por Russell en PCV *son insuficientes* ya que no justifican por sí mismas la inferencia de una contradicción. Muchos problemas que *no* tienen la *forma* de paradojas reflexivas responden a la descripción de PCV (véase 2.3). Pero tampoco Priest es más afortunado en este sentido, algunos razonamientos que no percibimos hoy en día como paradójicos *encajan* en EI: la "paradoja" del barbero [18] y el teorema de Cantor (véase 2.7) son dos buenos ejemplos. Así pues, no todas las contradicciones de inclusión –aquellos *razonamientos* que encajan en EI– son *paradojas* de inclusión (véase 2.8.2). La razón de este fracaso es quizá más profunda de lo que cabría suponer a primera vista. Tal vez los intentos fallidos de Priest por ofrecer una definición *estructural* de "paradoja reflexiva" no constituyan un fracaso personal sino una mera muestra de que tal tarea es *imposible*.

El mérito de EI consiste en identificar un tipo de razonamiento estrechamente vinculado con la formación de paradojas, pero existen al menos dos motivos importantes para poner en duda la posibilidad de ofrecer un análisis (estructural o de otro tipo) de las paradojas reflexivas en términos de condiciones necesarias y suficientes. Ambos motivos afectan a la posibilidad de ofrecer condiciones *suficientes*. El primer problema con que tropezamos en este sentido fue señalado por Priest en 5.3.1, aunque seguramente Kripke (1975) fue el primero en formularlo explícitamente (véase 4.5.1-2): En general, no existen criterios sintácticos, semánticos o estructurales que nos ayuden a determinar de antemano cuándo un enunciado es rechazable o paradójico y cuándo no.[1] Hemos de acudir siempre al contexto o al juicio que nos merecen los enunciados a los que apelamos en la demostración paradójica para decidir

[1] Como afirmamos en 5.3.1-2 existen *algunos* criterios sintácticos o estructurales (al menos si uno es un lógico consistente) que nos permiten determinar de antemano si un enunciado es rechazable o no. Por ejemplo, todo enunciado de la *forma* "$p \wedge \neg p$" es rechazable. Sin embargo, salvo en el caso citado, en el resto decidir si un enunciado es rechazable o no *no depende* de la presencia en dicho enunciado de uno u otro rasgo *sintáctico*. Para rechazar un enunciado, *p*, que no tiene la forma de una contradicción debemos examinar lo que dice en un contexto y evaluar la relación que mantiene con

si hemos *demostrado* una contradicción de hecho o, si habiendo demostrado una contradicción, ésta *parece aceptable* o es por el contrario tan rechazable como cualquiera de las que figuran en una *reductio*. El análisis estructural que ofrece El no proporciona razones que nos fuercen a aceptar la existencia de Ω o de δ. Sólo un examen detenido de la *interpretación* que Ω o δ reciben en cada caso nos ayudaría a determinar si realmente creemos en su existencia o en la posibilidad de definir $\delta(\Omega)$. Obviamente, ningún examen de este tipo podría alcanzar conclusiones generales válidas para *todas* las paradojas y ninguno de ellos tiene por qué basarse en criterios estructurales o sintácticos.

Este problema está relacionado con otro que ya señalamos en el capítulo 1. Si una paradoja tiene solución, la pregunta acerca de si la tiene o no se decidirá positivamente al encontrar dicha solución. Sin embargo, si una paradoja fuese "genuina" (irresoluble), entonces no podríamos demostrar que carece de solución. Tal conclusión sólo nos parecería aceptable y justificada si hubiésemos agotado antes *todas* las alternativas posibles. Ahora bien, ¿qué significa "agotar *todas* las alternativas"? ¿Cómo podemos saber si realmente las hemos agotado *todas* o si simplemente creemos (de forma errónea) haberlas agotado? No podemos saberlo. En general, la cuestión de si una paradoja es "genuina" resultaría indecidible si nos enfrentásemos a una que de hecho lo fuese. La historia de la humanidad muestra que muchos problemas que se consideraron irresolubles en su momento, entre ellos paradojas, se han resuelto con el

otros enunciados verdaderos. El resultado de este examen no está resumido en ningún rasgo sintáctico de p.

tiempo. Ahora bien, resolver una paradoja no consiste en *cambiar la forma* del razonamiento que la establece sino en invalidar dicho razonamiento o en incluirlo en una *reductio*. Si el razonamiento problemático se invalida y desaparece, no tiene sentido preguntar por su forma; si el razonamiento se reubica en una *reductio*, entonces *no cambia de forma*, a lo sumo le *añadimos* un paso más para inferir la negación de alguna de sus premisas, pero eso es todo. Esto significa que ningún análisis estructural puede determinar "qué es una paradoja", no podemos excluir la posibilidad de que haya razonamientos que parezcan paradójicos en un primer momento, que encajen en nuestro análisis y que, pese a ello, no sean paradójicos (i.e., no nos fuercen a aceptar la verdad de una contradicción). Este es el caso del teorema de Cantor y de la paradoja del barbero antes mencionados.

La definición de las paradojas de inclusión en términos de condiciones necesarias y suficientes se enfrenta también a un problema *formal*, sobre todo si un análisis de este tipo busca identificar cierta *totalidad*: la de todas las cosas que encajan en EI. Según el primer problema discutido, EI identifica la estructura de las *contradicciones* de inclusión y no la de las *paradojas* de inclusión, ya que no hay rasgos sintácticos que determinen cuándo algo es "paradójico" (mientras que la definición de "contradicción de inclusión" *sí* es sintáctica). El problema ahora es: ¿Podemos definir o determinar realmente el conjunto de *todas* las contradicciones de inclusión (sean o no paradójicas)? Suponer que esto es posible generará muy probablemente *paradojas*, cosa que no debería sorprendernos a estas alturas. Imaginemos que *CI* es el conjunto que agrupa a *todas* las contradicciones de inclusión (i.e., el conjunto de todas aquellas cosas que "encajan en EI"). Si consiguiésemos construir un diagonalizador, δ, que asignase a cualquier grupo X ⊆ *CI* de contradicciones de inclusión (o, al menos, a *CI*) un *nuevo* miembro de *CI*, entonces obtendríamos una contradicción de inclusión al considerar δ(*CI*), ya que δ(*CI*) *pertenecería y no pertenecería* a *CI*, esto es, δ(*CI*) *sería y no sería* una contradicción de inclusión. Construir un diagonalizador δ de estas características requeriría vencer ciertas dificultades técnicas (la principal sería idear un modo de describir sintácticamente la construcción de un nuevo miembro de *CI* a partir de aquellos miembros de X ⊆ *CI* sobre los que opera δ). Pero nada indica en principio que describir una operación de este tipo sea imposible y, si esto es así, la formación de una contradicción de inclusión resultará inevitable. Eso sí, esta contradicción de inclusión sería muy peculiar, ya que δ(*CI*) *es* una contradicción de inclusión si, y sólo si, *no es* una contradicción de inclusión.[2]

Así pues, vemos que cualquier intento encaminado a identificar un conjunto definido que englobe (*sólo*) a todo lo paradójico o incluso a todo lo que satisface el Esquema de Inclusión es en sí mismo

---

[2] Para construir un diagonalizador en el caso de *CI* probablemente deberíamos encontrar antes una manera estándar de codificar cualquier contradicción de inclusión y de ordenar los elementos de *CI* apoyándonos en dicha codificación. De este modo la construcción del diagonalizador δ y de δ(*CI*) sería más sencilla. Obtendríamos así un razonamiento que demuestra que δ(*CI*) es y no es una contradicción de inclusión, lo que probablemente nos forzaría a considerar soluciones consistentes como las descritas en el capítulo 4. Podríamos, por ejemplo, introducir una jerarquía de contradicciones de inclusión de distinto orden donde no existiera un único conjunto *CI*, sino varios: CI_0,

problemático y puede llegar a ser paradójico. Recapitulando lo dicho hasta ahora podemos identificar tres dificultades:

(1) Tenemos, por un lado, problemas relacionados con nuestra incapacidad para reconocer paradojas genuinas, estos problemas se basan en el rechazo *a priori* que suscita la aceptación de una paradoja y en la evidencia empírica de que algunas paradojas consideradas como irresolubles se han disuelto con el paso del tiempo. No encontramos, pues, justificable la actitud de alguien que acepta *hoy* el fracaso *absoluto* de la ciencia ante las paradojas. Esto hace que la tarea de identificar paradojas *genuinas* sea una misión imposible.

(2) Por otro lado, tenemos un problema de orden empírico (seguramente relacionado con el anterior) que nos impide determinar, sobre la base exclusiva de criterios sintácticos o estructurales, cuándo un razonamiento es paradójico (i.e., *percibido* como la justificación *aceptable* de una *falsedad*) y cuándo no lo es. Podemos encontrar contraejemplos a cualquier análisis existente (estructural o no) en términos de condiciones necesarias y suficientes de "lo paradójico", esto es, podemos encontrar razonamientos que *encajen* en el análisis propuesto pero que *no sean paradójicos*.

(3) Por último, tenemos un problema de "reflexividad", por así decirlo. Todo intento existente de identificar condiciones necesarias y suficientes que determinen "lo paradójico" tiende a generar nuevas paradojas, i.e., enunciados que satisfacen dichas condiciones si, y sólo si, *no las satisfacen*.

CI_1, CI_2 ... Tendríamos así un ejemplo de parametrización (con todos los problemas que

Éste último problema está relacionado con la discusión que hemos sostenido durante la segunda parte de este trabajo (capítulos 4 y 5). Identificar un grupo de condiciones que permita determinar por completo las paradojas de un lenguaje y "desactivar" sus efectos no deseados (ya sea extirpándolas o aceptándolas como dialetheias) es una tarea que nadie ha realizado con éxito hasta el momento. La estrategia general seguida por todos aquellos que se han enfrentado a las paradojas podría ser descrita vagamente de la siguiente manera: Cada teoría ofrece un marco conceptual que nos permite identificar, en primer lugar, (si no los enunciados problemáticos sí al menos) *el fenómeno* que causa las paradojas: circularidad viciosa, oraciones no fundamentadas, contradicciones verdaderas, etcétera. Las paradojas se resuelven a continuación a través de nuevas distinciones teóricas destinadas a mostrar que las situaciones paradójicas esconden algún tipo de ambigüedad (ya sea semántica o lógica), algún tipo de riqueza de sentidos o de alternativas lógicas que no percibimos a simple vista y que está relacionada con el fenómeno que las causa según la teoría. Resolver una paradoja consiste, pues, en hacer explícitas estas ambigüedades mediante el diseño de un marco lógico o conceptual más amplio en el que los términos paradójicos dejen de tener el sentido que tenían en la paradoja o en el que los razonamientos paradójicos dejen de ser válidos (o sean válidos pero no "paradójicos" –i.e., inaceptables– como ocurre en el caso de las lógicas paraconsistentes).

---

ello acarrea).

La paradoja de Parménides se disuelve cuando nos damos cuenta de que el verbo 'ser' puede ser utilizado en muchos sentidos (Aristóteles); la del mentiroso cuando descubrimos que la extensión del predicado verdadero debe relativizarse a un lenguaje (Tarski), o a un contexto de uso (Burge, Barwise, Simmons, etc.), o cuando nos damos cuenta de que algunas oraciones no pueden ser ni verdaderas ni falsas (Kripke, McGee, etc.); las paradojas del infinito se resuelven con el descubrimiento de números ordinales transfinitos, de cardinales infinitos y con la separación de las nociones de ordinal y cardinal (Cantor); la de Russell con la distinción entre conjuntos y clases (Von Neumann); las paradojas de inclusión se resuelven cuando nos percatamos de que hay dialetheias y, por tanto, razonamientos con conclusiones verdaderas y falsas (Priest) ... Todas estas soluciones parecen seguir el patrón que describimos en 4.3.1 apoyándonos en EI. Por desgracia, todas ellas parecen sucumbir ante el mismo problema (véase 4.3.2 y las discusiones concretas de las teorías tratadas): el marco conceptual o lógico que establecen genera *nuevas* paradojas, esto es, genera conceptos que deshacen las antiguas paradojas pero que permiten perfilar nuevas totalidades (la de los enunciados *no* fundamentados, o la de las contradicciones de *inclusión*, o la de los enunciados *sólo* falsos, o la de todas las *clases*, etc.) que podemos trascender mediante un diagonalizador para formar *nuevas* paradojas. El último recurso que nos queda para resolver estas paradojas de nuevo cuño es *prohibir* o *impedir* la formulación de totalidades peligrosas. Pero, como no podemos determinar sintáctica o estructuralmente las condiciones que determinan "lo paradójico", toda prohibición corre el riesgo de ser insuficiente o excesiva (lo que la convertiría en insatisfactoria en cualquier caso).

Estas consideraciones nos devuelven a los tres problemas anteriormente esbozados y a la vigencia de la conclusión que apoyaban: No tenemos (hasta la fecha) un método que nos permita *identificar* a *todas* las paradojas reflexivas y *sólo* a éstas. Decíamos en 4.2 que toda teoría preocupada por las paradojas debía satisfacer dos tareas fundamentales: la tarea de ofrecer un "diagnóstico" del problema y la tarea de ofrecer una "terapia" o solución ajustada al diagnóstico. Ahora podemos decir que ambas tareas están estrechamente relacionadas y que el fracaso "terapéutico" obedece en gran medida a las deficiencias en los "diagnósticos". Es muy difícil identificar un problema de forma precisa cuando no somos capaces de identificar a quienes lo padecen y si no hemos identificado el problema, "solucionarlo" es ya una quimera. Si las paradojas nos inquietan, es precisamente porque generan la sospecha de que, tal vez, nunca seamos capaces de identificar las condiciones *exactas* (i.e., necesarias *y suficientes*) que las provocan ni, por consiguiente, de protegernos contra ellas de forma segura y definitiva.

Ni Russell ni Priest han conseguido ofrecer condiciones necesarias y suficientes para identificar todas las "paradojas reflexivas", así pues, ni PCV ni EI pueden motivar una solución *uniforme a todas ellas y sólo a ellas* (como defendíamos en 2.3 y 5.3.6). Esa supuesta solución "uniforme" afectaría por igual a paradojas y a no paradojas y, si es agresiva, podría causar más daños que bienes. Pero entonces, ¿cuál es el valor real de EI? El que señalamos antes: La posibilidad de identificar una estructura abstracta que se manifiesta en la inmensa mayoría de las paradojas reflexivas (salvo en algunas versiones de la paradoja de Curry [17] sin negación, véase 2.8.1). EI nos ha permitido identificar un problema general y encontrar conexiones entre fenómenos aparentemente dispersos. Nos ha permitido también describir estrategias comunes de aproximación a las paradojas (4.3.1) y describir los motivos por los que esas estrategias suelen fracasar habitualmente (4.3.2). Así mismo, EI ofrece la posibilidad de establecer resultados generales basados en la estructura de las contradicciones de inclusión que pueden extenderse a todos los razonamientos que encajan en EI sin importar dónde se produzcan o qué conceptos involucren. Por otra parte, el análisis estructural que ofrece EI de las contradicciones de inclusión puede complementarse con análisis estructurales de fenómenos relacionados con éstas, como, por ejemplo, la circularidad viciosa (presente en Ω a través de uno de sus miembros $\delta(\Omega)$),[3] o la descripción de estrategias para construir argumentos diagonales con δ como elemento clave.

[3] En este sentido la noción semántica de fundamentación [*groundedness*], a la que han apelado tanto Herzerberger (1970a) como Kripke (1975) en sus reflexiones sobre paradojas puede ser muy útil a la hora de describir fenómenos de circularidad viciosa en relación con ciertos proyectos semánticos (i.e., ofrecer una definición de la noción de verdad), lógicos y epistemológicos (i.e., establecer la validez de los principios que rigen el pensamiento o la posibilidad del conocimiento). Herzerberger y Kripke definen '*groundedness*' de diferente manera pero en ambos casos lo que persiguen (entre otras cosas) es captar y describir estructuralmente un fenómeno de circularidad viciosa que está presente también en la noción de conjunto no fundamentado [*non-wellfounded set*] trabajada en nuestros días por autores como Aczel (1988) o Barwise y Moss (1996). En J. Valor (2001) esbozo algunas ideas relativas a una posible reformulación del término "groundedness" (distinta a las de Herzerberger y Kripke) que tendría por objetivo dar un uso más amplio a este término. Se trataría de presentar una acción como algo que procede mediante la ejecución de una serie de reglas que fundamentan su éxito final. La ejecución de algunas de estas reglas (que cabría describir mecánicamente) podría

Es cierto que hablar de la *totalidad* de las contradicciones de inclusión podría generar paradojas, quizá tal *totalidad* sea inconsistente pero, ¿significa esto que la categoría "contradicción de inclusión" es inútil? No lo creo. Conceptos como "verdad" o "conjunto" son inconsistentes cuando consideramos la totalidad de las verdades o la de los conjuntos, ¿significa eso que debemos abandonar estos conceptos? No. Los conceptos son útiles, nos ayudan a pensar y a representar una cantidad ingente de información con la que no podríamos operar de otro modo. La noción de "contradicción de inclusión" es útil en el mismo sentido. Nos ayuda a introducir una unidad no forzada o artificial en fenómenos que, de hecho, comparten similitudes estructurales relevantes. Ciertamente, su utilidad no elimina las paradojas, como tampoco las elimina en el caso de las nociones de verdad o de conjunto, pero *ése* es el problema al que nos enfrentamos cuando nos acercamos a las paradojas, a saber, la imposibilidad de ofrecer teorías generales consistentes sobre conceptos fundamentales.

La situación a la que llegamos en este caso es chocante y (¡cómo no!) paradójica. Tal vez no podamos hablar coherentemente acerca de *todas* las contradicciones de inclusión pero, si tropezamos con una paradoja reflexiva (por ejemplo la que establece que $\delta(CI)$ *es y no es* una contradicción de inclusión), es más que probable que *encaje* en el EI y, si esto es así, podremos "mostrar" de algún modo que EI consigue describir en buena medida el fenómeno relevante. Esto provocaría una situación

presuponer en ciertas circunstancias (no determinables *a priori*) el cumplimiento de condiciones circulares, en cuyo caso la acción que dichas reglas describen no estaría bien fundamentada.

curiosa, afirmar que $\delta(CI)$ –la posible paradoja originada por EI– es una contradicción de inclusión nos llevaría a inferir una contradicción, sin embargo, mostrar que $\delta(CI)$ encaja en EI resultaría tan simple como examinar la forma que tiene de hecho la inferencia de la contradicción, tenemos así una paradoja que *encaja* en EI, pero de la que sólo podemos *afirmar que encaja* en EI si, y sólo si, no encaja en EI. De algún modo, y parafraseando a Wittgenstein, podríamos "*mostrar*" que $\delta(CI)$ es una contradicción de inclusión, pero no podríamos "*decir*" coherentemente que lo es.

6.2 De nuevo el viejo problema

Esto nos devuelve al problema con el que iniciamos este trabajo. Las paradojas nos parecen en general heridas abiertas que no acaban de cicatrizar. Después de muchos siglos aún nos acompañan muchas de ellas y todo intento de solución ha desembocado en la formación de nuevas paradojas (a menudo versiones reforzadas de las anteriores). Es más, todo esfuerzo encaminado a *delimitar* el ámbito de "lo paradójico" ha tropezado con problemas de coherencia provocados precisamente por la formación de paradojas. El Esquema de Inclusión no cambia esta situación. Cualquier "diagnóstico" o "terapia" relativos a las paradojas que queramos basar en EI tendrá límites. EI (al igual que otros análisis) no consigue ofrecer condiciones necesarias y suficientes para determinar *sólo* a las *paradojas* reflexivas. Pero, además, EI no consigue delimitar coherentemente el conjunto de todas las *contradicciones* de inclusión, ya que (como también ocurría con el análisis de otros términos: "verdad", "conjunto", etc.) suponer que podemos hablar de la *totalidad* de las cosas que encajan en EI genera probablemente paradojas de inclusión.

Así pues, ¿cuál es el estatus de un discurso que toma a las paradojas por objeto? Hoy por hoy la ciencia sólo puede convivir con ellas si consigue destruirlas. Un discurso científico que se ocupe de las paradojas sólo puede tener dos aspiraciones: *explicarlas* y *resolverlas*. Tal vez no encuentre un modo obvio de alcanzar su objetivo a corto plazo, pero no puede cejar en su empeño. No hay conocimiento de lo *falso*, por tanto ningún razonamiento que no tenga por finalidad la verdad puede ser válido o aceptable. Todo discurso científico que uno pueda construir sobre las paradojas ha de perseguir su eliminación deshaciendo la *ilusión* que encierran. Si los sucesivos intentos por resolver ciertas paradojas se muestran infructuosos en un área determinada, la ciencia puede optar en último extremo por excluir dicha área de su ámbito de acción, aunque para ello *deberá mostrar* que la inferencia de falsedades es realmente *inevitable* en esa área y que es satisfactorio abandonar su estudio. Si esto no es posible, lo único que puede hacer es seguir insistiendo en sus esfuerzos por resolver la paradoja aferrándose a la esperanza de que algún día ésta se disuelva. Claro que, si la paradoja no tiene solución y es imposible demostrar este hecho, sus esfuerzos serán vanos y no nos acercarán en última instancia a la verdad.

La filosofía, por su parte, se encuentra en el mismo dilema que cualquier otra disciplina encaminada a obtener conocimiento cuando se acerca a las paradojas. Es imposible conceder validez a un razonamiento que justifique falsedades y, por tanto, las paradojas deben rechazarse. Sin embargo, el problema de las paradojas cobra aquí un protagonismo especial. Ciertamente, no disponemos de una solución satisfactoria para ellas, pero nuestra experiencia nos dice que, en la mayoría de los casos, podemos ignorarlas, y no sólo en nuestra vida cotidiana sino también en las prácticas científicas. Las paradojas sólo suponen un obstáculo para la ciencia cuando ésta se enfrenta a ciertas cuestiones básicas de carácter fundacional como definir lo que entendemos por "verdad", "conjunto", significado", "conocimiento", etcétera. Aunque la ciencia no disponga de teorías plenamente satisfactorias con respecto a estos conceptos elementales, sí dispone al menos de teorías que nos permiten usarlos en la mayoría de casos sin incurrir en contradicciones. De ese modo las tareas de ofrecer definiciones generales para conceptos básicos y respuestas a cuestiones fundacionales pueden aplazarse sin más problemas. La filosofía, en cambio, no puede dejar de reflexionar sobre las paradojas y sobre aquellos ámbitos en los que habitualmente tropezamos con ellas. Y, al hacerlo, no hace sino generar nuevas paradojas y perplejidades. Al intentar entender cuál es el sentido último de ciertos términos y conceptos relacionados con nuestra vida cognitiva, la filosofía se ha preguntado a lo largo de su historia cuáles son los límites de lo pensable, de lo cognoscible, de lo racional, los límites del yo, del sentido, etc., buscando a menudo un modo de definir qué entendemos por conocimiento, razón, sujeto, significado, etc. y un modo de establecer si es posible la existencia de dichas entidades. Sin embargo, la idea de que estas preguntas puedan responderse mediante una investigación empírica de carácter científico es problemática. De hecho, una de las cosas que está en juego es entender *si son posibles la ciencia y el conocimiento empírico y qué son.* Llevar a cabo una investigación *científica* para responder a esta pregunta es *haberla respondido ya de antemano.* Toda investigación que iniciemos sobre las nociones de sujeto, razón, ciencia, conocimiento o significado estará condicionada por lo que intentamos establecer. Ninguna experiencia que examinemos *podría entenderse* sin ser ubicada en un "*yo*" que *razona* de acuerdo con su constitución a partir de lo que *conoce* y mediante un lenguaje cuyo significado *entiende.*

La filosofía se enfrenta pues a un problema grave de *circularidad viciosa*: Si es posible construir una ciencia que responda las preguntas que preocupan a la filosofía, esa ciencia deberá justificarse a sí misma, lo que resulta tan sospechoso como organizar un juicio en el que los papeles de acusado y juez recaen sobre la misma persona (y en el que el "juez" sólo puede juzgar al "acusado" cometiendo los delitos de los que le acusa). Sin embargo, si la filosofía desecha la validez de esta posible ciencia (a la que tal vez cabría llamar metafísica) e intenta justificar su rechazo, se enfrenta a una *paradoja de inclusión*: Si ninguna ciencia puede responder estas preguntas, entonces la filosofía no puede tener estatus de ciencia cuando aborda su respuesta tentativamente. Ahora bien, una manera de *responder* estas preguntas (i.e., de cancelar el problema que plantean) es afirmar que "no tienen respuesta" y añadir que, en la medida en que la filosofía no puede abandonar la tarea de responderlas, "la filosofía no es ciencia" ya que no aporta conocimiento sino "otra cosa". Pero, ¿qué estatus tienen estas afirmaciones hechas desde la filosofía? No pueden tener un estatus científico porque se autorrefutarían; pero, si no tienen un estatus científico, son especulaciones que podrían resultar falsas, lo que reintroduce la posibilidad de que exista una respuesta rigurosa y "científica" a esas preguntas (posibilidad cuya realización tropezaba, no obstante, con problemas de circularidad). Por otra parte, cuando la filosofía convierte esta paradoja interna en objeto de reflexión, volvemos a planteamos un problema sobre límites ("¿cuál es el límite de las afirmaciones filosóficas?") que a su vez requiere responder las preguntas anteriores y que es víctima por ello de nuevas paradojas.

Los problemas de coherencia y de circularidad que generan las paradojas parecen no tener fin. Tanto si abordamos paradojas básicas (i.e., la del mentiroso, la de Russell, etc.) como si abordamos paradojas de orden superior (i.e., aquellas que surgen cuando reflexionamos sobre la significación de las paradojas en nuestra vida cognitiva), tropezamos siempre con contradicciones y paradojas "reforzadas". Quizá lo más constante en todas ellas (y, por ello, el objeto de estudio más seguro) sea su forma: la presencia de cierto tipo de circularidad viciosa en la definición de una totalidad, X, y de uno de sus elementos, *a*, y la forma en que dicho elemento es construido mediante una operación que, por su naturaleza, parece situar a *a* dentro y fuera de X a la vez. Pero, sea como fuere, ninguna descripción formal puede cancelar o resolver el problema que suponen las paradojas. Y, además, hay motivos para pensar que ni siquiera podría ofrecer condiciones formales necesarias *y suficientes* para determinar el ámbito de lo paradójico. Así pues, el problema de fondo sigue en pie, algunos razonamientos parecen (inaceptablemente) justificar falsedades y no sabemos muy bien cómo evitar este hecho sin renunciar a principios que nos parecen incuestionables. El problema que presentan las paradojas no es "formal", no responde al cumplimiento de cierto patrón estructural por parte de uno u otro razonamiento, sino a nuestras creencias con respecto a la verdad de ciertos enunciados y principios y a la visión del mundo que se desprende de dichas creencias.

APÉNDICE

PARADOJA (del Latín 'paradoxus' y éste del griego 'παραδοξος'):

I. *Diccionario de la Real Academia Española de la lengua*, vigésima segunda edición, 2001 (DRAE):
(1) Idea extraña u opuesta a la común opinión y al sentir de las personas; (2) Aserción inverosímil o absurda que se presenta con apariencia de verdadera; (3) Figura de pensamiento que consiste en emplear expresiones o frases que envuelven contradicción: *mira el avaro en sus riquezas pobre*.

II. *Diccionari de la llengua catalana de l'Institut d'Estudis Catalans*, segona edició, 1995 (DIEC):
(1) Opinió contrària a l'opinió comuna, especialment que sembla contrària al comú sentir però que de fet és exacta; (2) En lògica, enunciat o raonament que porta a dues conclusions mútuament contradictòries però de cap de les quals no es pot prescindir; (3) Figura retòrica per la qual un enunciat veritable, però que sol constar de dos conceptes antagònics, és presentat d'una manera inversemblant, absurda o contradictòria.

III *The new shorter Oxford English dictionary on historical principles*, (Ed. Lesley Brown), Oxford UP, Oxford, 1993.
(1) A statement or tenet contrary to received opinion or belief, especially one that is incredible, absurd or fantastic. (2a) A seemingly absurd or self-contradictory statement or proposition which when investigated or explained may prove to be well-founded or true. (2b) A proposition or statement that is actually self-contradictory, absurd or false. 2c) Logic. In full "logical paradox". A statement or proposition which, despite sound reasoning from an acceptable premiss, leads to a conclusion that is against sense, logically unacceptable or self-contradictory. (3) Paradoxical character, condition or quality. (4) A phenomenon that exhibits some contradiction or conflict with preconceived notions of what is reasonable or possible; (5) a person of perplexingly inconsistent life or behaviour.

ÍNDICE DE PARADOJAS REFLEXIVAS

[1] ***La paradoja del mentiroso.*** Su versión más popular es el enunciado "Estoy mintiendo", que es verdadero si y sólo si es falso. La versión histórica de esta paradoja se encuentra en la Epístola de San Pablo a los Corintios donde se nos habla de un cretense, Epiménides, que afirma "Todos los cretenses mienten siempre". Estrictamente hablando, este enunciado no es necesariamente paradójico, ya que podría ser falso. Sin embargo, si ningún cretense, incluido Epiménides, ha dicho nunca la verdad, tenemos una paradoja. Se llama "mentirosos contingentes" a las oraciones que pueden convertirse en paradojas dependiendo de su contexto de uso. A lo largo del presente trabajo examinaremos diferentes versiones de la paradoja del mentiroso. En todas ellas encontramos oraciones (enunciados o proposiciones) que son verdaderas si y sólo si son falsas. Las versiones más populares del mentiroso son: μ = 'μ es falsa', μ^* = 'μ^* no es verdadera' (mentiroso reforzado: no puede ser ni verdadera, ni falsa, ni carente de valor de verdad coherentemente). Por último encontramos las cadenas de mentirosos: μ_0 = 'μ_1 es verdadera'; μ_1 = 'μ_0 es falsa' (de longitud variable).

[2] ***La paradoja de Russell.*** Afecta a la clase de todas aquellas clases que no se pertenecen a sí mismas. Dada su definición, dicha clase se pertenece a sí misma si y sólo si no se pertenece a sí misma.

[3] ***La paradoja relacional.*** Construida a partir de la relación T que mantienen dos relaciones cualesquiera R y S entre sí cuando ocurre que R *no mantiene* la relación R con S, esto es, $T(R,S) \leftrightarrow \neg R(R,S)$. El problema radica en que $T(T,S) \leftrightarrow \neg T(T,S)$.

[4] ***La paradoja de Berry.*** Pensemos en "el1 me2nor3 nú4me5ro6 en7te8ro9 que10 no11 se12 pue13de14 nom15brar16 con17 me18nos19 de20 vein21ti22o23cho24 sí25la26bas27". Si el número existe, no se

puede nombrar con menos de veintiocho sílabas, sin embargo la descripción anterior lo nombra y tiene menos de veintiocho sílabas.

[5] ***La paradoja de König***. Hace referencia al menor número ordinal indefinible (o indescriptible). Tal número existe dado que el conjunto total de posibles definiciones (o descripciones) en castellano es enumerable, mientras que el conjunto de los ordinales no lo es y existen, por tanto, más ordinales de los que podemos describir o definir. Por otra parte, todo conjunto de números ordinales tiene siempre un elemento menor, por lo que el conjunto de números ordinales indefinibles también ha de tenerlo, pero "el menor número ordinal indefinible (o indescriptible)" es una definición (o descripción) del citado ordinal indefinible.

[6] ***La paradoja de Richard***. Versa sobre el conjunto de los números reales entre 0 y 1 (con expansiones decimales infinitas) definibles mediante un número finito de palabras, por ejemplo: "cero, seguido de infinitos unos". El conjunto de estas definiciones es enumerable, mientras que el citado subconjunto de números reales no lo es. Así pues, podemos asignar un número natural a cada definición y ofrecer una lista numerada, L, de todas ellas. Consideremos ahora el siguiente decimal infinito: D = "Cero seguido del primer decimal del primer número de la lista L, del segundo decimal del segundo número de L y, para cualquier posición decimal n, del enésimo decimal del enésimo número de L". Consideremos ahora a D' = El número resultante de sustituir n por n+1 (para todo número natural n) en todas las posiciones decimales de D. D' está en la lista L si y sólo si no lo está.

[7] ***La paradoja de Burali-Forti***. Afecta a la existencia del mayor ordinal posible o, alternativamente, de la clase de todos los ordinales: 1) A todo conjunto bien ordenado le corresponde un ordinal. 2) Un ordinal α es igual al conjunto de todos los ordinales que le preceden menos α ($\alpha \notin \alpha$). 3) Todo ordinal α tiene un sucesor distinto de él, $\alpha \cup \{\alpha\} \neq \alpha$. 4) La colección de todos los ordinales está bien ordenada (de menor a mayor) y,

por tanto, ha de corresponderle un *ordinal*, a saber, la clase de todos los ordinales: On (que es ella misma un ordinal). Ahora bien, si On es un ordinal, su sucesor, On $\cup$ {On}, debe ser también un ordinal distinto de él. Pero eso no es posible porque On es la clase de todos los ordinales y, por tanto, On $\in$ On, con lo que On $\subseteq$ On $\cup$ {On} y On $\cup$ {On} $\subseteq$ On, esto es, On = On $\cup$ {On}.

[8] ***La paradoja de Grelling/Weyl.*** Concierne al predicado 'heterológico', predicado verdadero de todos aquellos predicados que *no* son verdaderos de sí mismos. 'Verde', por ejemplo, es heterológico, ya que no es verde, 'corto' en cambio no lo es, porque es corto. Pero 'heterológico' es heterológico si y sólo si no lo es.

[9] ***La paradoja de Mirimanoff.*** O, también, de la jerarquía bien fundamentada de conjuntos. Partiendo del conjunto vacío, $\varnothing$, podemos generar una secuencia transfinita de conjuntos: $\varnothing$, $\wp(\varnothing)$, $\wp(\wp(\varnothing))$, ... en la que cada elemento sucesor es el conjunto potencia de su antecesor: $x_{\alpha+1} = \wp(x_\alpha)$, y cada elemento límite, la unión de sus antecesores: $x_\lambda = \bigcup\{x_\alpha: \alpha < \lambda\}$ (α, $\lambda \in$ On, λ es un ordinal límite). La *unión* de todos los elementos de esta serie recibe el nombre de "Jerarquía Acumulativa" ("R" para abreviar). R tiene tres propiedades importantes:

(1) Todo miembro de R es un subconjunto *propio* de cualquier otro que le suceda en la secuencia y, también, un subconjunto de R ($\forall a, a \in \mathrm{R} \rightarrow a \subseteq \mathrm{R}$).

(2) Todos los elementos de R son conjuntos *fundamentados*. (a_0 está fundamentado cuando es imposible construir a partir de él una serie infinita descendente –i.e., *un regreso al infinito*– mediante la relación de pertenencia: $a_0 \ni a_1 \ni a_2 \ni$ ... Toda secuencia descendente acaba forzosamente en $\varnothing$ tras un número *finito* de pasos.).

(3) R es el conjunto de *todos* los conjuntos bien fundamentados.

La paradoja radica en que, si R está fundamentado, entonces, por (3), R $\in$ R, lo que implica que no está fundamentado (ya que R $\ni$ R $\ni$ R $\ni$ ...); pero si R no está fundamentado, entonces, por (2), R $\notin$ R lo que implica que R está fundamentado (pues todos sus miembros lo están).

[10] ***La paradoja de Cantor.*** Consideremos la clase de todos los conjuntos puros, V, formemos ahora la clase de todos sus subconjuntos, $\wp(V)$. Todos los elementos de $\wp(V)$ son conjuntos, por tanto $\wp(V) \subseteq V$, ahora bien, todos los elementos de V son *subconjuntos* de V (conjuntos de conjuntos que pertenecen a V), por lo tanto $V \subseteq \wp(V)$, así pues, $\wp(V) = V$. Pero esto es imposible por el teorema de Cantor [I].

[11] ***La paradoja de la cita (Tarski).*** Si entendemos las comillas como funciones que asignan a cada término lingüístico un nombre del mismo al entrecomillarlo, obtenemos una paradoja semántica que prescinde del predicado 'verdadero'. Pensemos en un lenguaje en el que podamos construir la siguiente oración autorreferente: c = "$\forall$p (c = 'p' $\rightarrow \neg$ p)" (i.e., 'c' es un nombre de la oración "$\forall$p (c = 'p' $\rightarrow \neg$ p)"). Si (para cualesquiera dos oraciones P y Q) se cumple que 'P' = 'Q' $\rightarrow$ (P $\leftrightarrow$ Q) es un principio válido (como parece serlo si las comillas son, de hecho, una función), obtenemos una paradoja, i.e., una contradicción demostrada. Aquí está el razonamiento[1]:

1. c = '$\forall$p (c = 'p' $\rightarrow \neg$ p)'	(Premisa 1)
2. 'P' = 'Q' $\rightarrow$ (P $\leftrightarrow$ Q)	(Premisa 2)
3. $\forall$p (c = 'p' $\rightarrow \neg$ p)	(Suposición)
4. c = '$\forall$p (c = 'p' $\rightarrow \neg$ p)' $\rightarrow \neg$ $\forall$p (c = 'p' $\rightarrow \neg$ p)	(instancia de 3)
5. $\neg$ $\forall$p (c = 'p' $\rightarrow \neg$ p)	(MP: 1, 4)
6. $\neg$ $\forall$p (c = 'p' $\rightarrow \neg$ p)	(absurdo: 3-5)
7. $\exists$p $\neg$ (c = 'p' $\rightarrow \neg$ p)	(Interdef. $\forall$, $\exists$: 6)
8. $\exists$p (c = 'p' $\wedge$ p)	(Interdef. $\rightarrow$, $\wedge$: 7)
9. c = 'q' $\wedge$ q	(Sup. -prueba por casos $\exists$ en 8)
10. 'q' = c = '$\forall$p (c = 'p' $\rightarrow \neg$ p)'	(Transitiv. de '=': 1, 9)
11. q $\leftrightarrow$ $\forall$p (c = 'p' $\rightarrow \neg$ p)	(2, 10)
12. $\forall$p (c = 'p' $\rightarrow \neg$ p)	(Mod. Pon: 9 (q), 11 (q $\rightarrow$ 12))
13. $\forall$p (c = 'p' $\rightarrow \neg$ p)	(Elimina $\exists$: 9-12)
14. Contradicción: 13, 6	

[1] Las suposiciones y sus cancelaciones se representan aquí mediante tabulaciones.

[12] ***La paradoja de Barwise y Moss***. Partiendo de un lenguaje, *L*, cuyos términos denotativos sólo pueden tener *un* referente en cada contexto de evaluación, Barwise y Moss consideran el término denotativo, τ, cuya interpretación se define circularmente del siguiente modo. Dado un contexto de evaluación:

(1) τ designa a 1 si τ designa a 0.
(2) τ designa a 0 si τ designa a 1.
(3) τ designa a 1 si τ carece de referencia.

Si τ designa a 1, entonces designa a 0, lo cual es imposible. Si τ designa a 0, entonces designa a 1, lo que también es imposible. Por tanto τ no puede tener referencia, ahora bien, si ése es el caso, τ designa a 1. Todos los supuestos conducen a una contradicción.

[13] ***La paradoja de Yablo***. Se formula mediante una serie infinita de oraciones s_0, s_1, s_2, ... en la que cada oración de la serie dice de sus sucesoras que son falsas:

(s_0) Para todo $k > 0$, s_k no es verdadera.
(s_1) Para todo $k > 1$, s_k no es verdadera.
(s_2) Para todo $k > 2$, s_k no es verdadera. ...

Supongamos (para *cualquier* n) que s_n es verdadera, entonces para *todo* $m > n$, s_m es falsa. Tomemos s_{n+1}, por ejemplo: s_{n+1} ha de ser falsa (dado que $n+1 > n$), pero si s_{n+1} es falsa, entonces debe existir un $k > n+1$ tal que s_k sea verdadera. Sin embargo, esto contradice la verdad de s_n ya que $k > n$. Ahora bien, si suponemos que s_n es falsa, entonces debe existir una oración verdadera s_k, con $k > n$, pero como acabamos de ver, la verdad de cualquier oración de la serie implica una contradicción. La peculiaridad de esta serie de oraciones es que consigue generar una paradoja que involucra al predicado 'verdadero' prescindiendo, aparentemente, de cualquier forma de autorreferencia o circularidad. Ninguna oración habla de sí misma o de sus predecesoras en la serie.

[14] ***La paradoja del cocodrilo***. Un niño cae al Nilo y es capturado por un cocodrilo, el padre del niño implora al cocodrilo que le devuelva a su hijo. El cocodrilo le promete devolvérselo sólo si adivina qué va a hacer con él (y comérselo en caso contrario). Tras pensar un momento, el padre

responde: “te lo comerás”, con lo que el cocodrilo puede cumplir su promesa si y sólo si no la cumple.

[15] ***La paradoja del Hiper-juego***. En algunos juegos para dos personas donde A mueve en primer lugar y B en segundo, existe siempre una estrategia ganadora para uno de los dos jugadores desde el principio de cualquier partida. Llamemos “cerrados” a estos juegos. El Hiper-juego es un juego para dos personas con las siguientes reglas básicas: 1) El primer movimiento de toda partida consiste en la elección, por parte de A, de un juego cerrado, J. 2) En el segundo movimiento B ocupa la posición de jugador A (primer jugador) en J y realiza un movimiento en dicho juego. 3) El Hiper-juego continua como una partida normal del juego J hasta su final. El Hiper-juego debe ser un juego cerrado, ya que cualquier juego que se elija en el primer movimiento lo es. Pero, si el Hiper-juego es un juego cerrado, entonces es elegible por A en su primer movimiento, y después por B, y por A de nuevo, ... Tendríamos así una partida abierta, sin ganadores, y, consiguientemente, el Hiper-juego no sería un juego cerrado. La paradoja consiste, pues, en que el Hiper-juego es un juego cerrado si y sólo si no lo es.

[16] ***La paradoja del conocedor*** [*The knower's paradox*]. Se trata de un esquema en el que encaja una familia de paradojas de carácter cognitivo. Convengamos en llamar respectivamente *Esquema de la Cognición* y *Esquema Negativo de la Cognición* a:

(EC) $\forall x\,(x \in \Sigma \rightarrow x \in \mathrm{C})$ (ENC) $\forall x\,(x \in \Sigma \rightarrow x \notin \mathrm{C})$

donde x es una oración y C es la extensión de un predicado cognitivo (‘cognoscible’, ‘verdadero’, ‘demostrable’, etc.) *cuya satisfacción implica ser verdadero*: $x \in \mathrm{C} \rightarrow x \in \mathrm{V}$ (V es el conjunto de las verdades). Intuitivamente, los esquemas dicen: “Si x es una oración de cierto tipo (del tipo que marca la propiedad que define Σ), entonces x (no) es cognoscible o verdadera o ...” La paradoja surge, a partir de ENC, cuando tomamos una oración φ que satisface dos condiciones: (1) $\varphi = \forall x\,(x \in \Sigma \rightarrow x \notin \mathrm{C})$, i.e., dada una interpretación de Σ y C, φ es una *ejemplificación* de ENC; (2) $\underline{\varphi} \in \Sigma$, i.e., donde Σ es un conjunto arbitrario cuya única restricción consiste en que tiene, al menos, un miembro: φ (utilizo el subrayado para construir nombres: '$\underline{\varphi}$' nombra a la oración 'φ').

[Antes de ver la demostración en abstracto, consideremos una ejemplo concreto. Si interpretamos C como la propiedad de 'expresar un hecho cognoscible' y Σ como la propiedad de 'ser una oración significativa', ENC se convierte en una versión del Escepticismo General: EG = 'Si x es una oración significativa, no expresa un hecho que podamos conocer'.]

Veamos cómo se infiere la paradoja: Supongamos que $\varphi \in C$ ...

1. $\underline{\varphi} \in C$	(suposición)
2. $\underline{\varphi} \in V$	($\underline{\varphi} \in C \rightarrow \underline{\varphi} \in V$ por def. C: 1)
3. φ	(Esquema V tarsquiano ($\underline{\varphi} \in V \leftrightarrow \varphi$): 2)
4. $\forall x\,(x \in \Sigma \rightarrow x \notin C)$	(condición (1) def. de φ: 3)
5. $\underline{\varphi} \in \Sigma \rightarrow \underline{\varphi} \notin C$	(Eliminación $\forall$: 4)
6. $\underline{\varphi} \in \Sigma$	(condición (2) def. de φ)
7. $\underline{\varphi} \notin C$	(Modus Ponens: 5, 6)
8. $\underline{\varphi} \notin C$	(absurdo: 1-7 por contradicción 1,7)

Hasta aquí todo va relativamente "bien", nos hemos limitado a demostrar (bajo los supuestos especificados: φ y $\varphi \in \Sigma$) que φ no puede ser C coherentemente [bajo la interpretación propuesta, que la hipótesis del escéptico, EG, no puede ser conocida]. Esto no entraña ninguna contradicción o absurdo [aunque genera la misma perplejidad que el argumento del escéptico], sin embargo, si ofrecemos un argumento sólido (y de carácter sustancial, no meramente formal) a favor de la tesis $\varphi \in C$ [como por ejemplo el que nos ofrece el escéptico], entonces obtenemos una contradicción. Además de la paradoja del escéptico, también encaja en este esquema la paradoja del relativista.

[17] ***La paradoja de Curry** (o de Löb)*. Existen muchas variantes de esta paradoja, las más simples se construyen a partir de oraciones del tipo:

(s) Si s es verdadera, entonces p.

donde 'p' puede ser cualquier oración: 'Dios existe', 'Dios no existe', '2 = 3', ... La oración s dice de sí misma que si es verdadera, entonces es el caso que p. El problema que plantea *s* es que a partir de ella es posible demostrar p, es decir, cualquier cosa -incluyendo contradicciones o falsedades. Veamos como.

1. *s* es verdadera	(suposición)
2. *s* es verdadera → p	(Esq.-T: "x es verdadera ↔ p" + def. *s*: 1)
3. p	(Modus Ponens,: 1, 2)
4. *s* es verdadera → p	(Introducción →: 1-3)
5. *s* es verdadera	(Esquema-T: + def. *s*: 4)
6. p	(Modus Ponens: 4, 5)

Co[illegible]ualquier cosa, podemos demostrar un absurdo. Po[illegible] del siguiente modo: s = '$s \in T \rightarrow \perp$', donde T es la colección de todas las verdades y $\perp$ una falsedad lógica. La paradoja de Curry muestra que a partir de *s* podemos demostrar $\perp$.

[18] ***La paradoja del barbero***. Se trata aquí de imaginar a un barbero que afeita a todas aquellas personas que no se afeitan a sí mismas. La paradoja surge cuando nos percatamos de que ese barbero se afeita a sí mismo si y sólo si no se afeita a sí mismo.

[19] ***La "Quinta Antinomia"*** (Priest, *BLT*, p. 100). Ésta es una paradoja basada en las antinomias kantianas. Consideremos el siguiente operador iterativo: 'El pensamiento que versa sobre *x*' (p(x) para abreviar). Supongamos que generamos una secuencia transfinita de pensamientos aplicando reiteradamente este operador a partir de un objeto inicial, por ejemplo, "la crítica de la razón pura": x_0 = la crítica de la razón pura; x_1 = p(la crítica de la razón pura); x_2 = p(p(la crítica de la razón pura)); ... Cuando llegamos a una secuencia infinita de pensamientos, los agrupamos a todos: X = $\{x_0, x_1, x_2, ...\}$, y continuamos la serie con un elemento límite aplicando de nuevo el operador 'p': x_ω = p(X); $x_{\omega+1}$ = p(p(X)); ... Como se puede apreciar, ningún pensamiento en la serie así formada puede ser reflexivo o autorreferente, ya que la definición de cualquier *x* en la secuencia es independiente de (y previa a) p(x). Si pensamos en el límite del operador 'p', es decir, si pensamos en la

colección P que reúne a todos los pensamientos formados mediante la iteración de 'el pensamiento de' (sobre un objeto inicial x), entonces nos enfrentamos a una paradoja. Por un lado, al pensar en P, mostramos que es posible formar 'el pensamiento que versa sobre P' que aparece, además, como un elemento adicional de la secuencia (un elemento límite de la misma); por otro, ningún pensamiento así formado puede quedar fuera de P, por tanto 'el pensamiento que versa sobre P' es reflexivo (i.e., pertenece a P), lo que, como vimos, es imposible.

Para más referencias sobre paradojas consúltese Russell 1908, Sainsbury 1995 o Priest 2002.

BIBLIOGRAFIA

ACZEL, P., 1988, *Non-Well-Founded Sets*, CSLI Publications, Stanford.

ARMOUR-GARB, B., 2001, "Deflationism and the Meaningless Strategy", *Analysis* 61: 280-89.

ARMOUR-GARB, B. & BEALL, J.C., 2001, "Can Deflationists Be Dialetheists?", *Journal of Philosophical Logic* 30: 593-608.

ARMOUR-GARB, B. & BEALL, J.C., 2002, "Further Remarks on Truth and Contradiction", *Philosophical Quarterly* 52: 217-25.

ANDERSON, A. R., 1970, "St. Paul's Epistle to Titus", en MARTIN 1978, pp. 1-11.

ARISTÓTELES, 1982, *Tratados de lógica (Organon) I: Categorías- Tópicos- Sobre las refutaciones sofísticas*, traducción al castellano de Miguel Candel Sanmartín, Gredos, Madrid.

ARISTÓTELES, 1994, *Metafísica*, traducción al castellano de Tomás Calvo Martínez, Gredos, Madrid.

AUSTIN, J. L., 1950, "Truth", *Proceedings of the Aristotelian Society, suppl. vol.* 24: 111-29. Reimpreso en AUSTIN 1979, pp. 117-33 (referencias a esta reimpresión).

AUSTIN, J. L., 1979, *Philosophical Papers* (eds. J. O. Urmson & G. J. Warnock), Oxford UP, Oxford, 3ª edición (1ª edición 1961). Traducción al castellano de Alfonso García Suárez: *Ensayos filosóficos*, Alianza Editorial, Madrid, 1989.

BARBA, J., 1998, "Construction of Truth Predicates: Approximation versus Revision". *Bulletin of Symbolic Logic* 4: 399-417.

BARTLETT, S. J. & SUBER, P., 1987, *Self-Reference. Reflections on Reflexivity*, M. Nijhoff, Dordrecht.

BARWISE, J., & ETCHEMENDY, J., 1987, *The Liar. An Essay on Truth and Circularity*, Oxford UP, Oxford.

BARWISE, J., & MOSS, L., 1996, *Vicious Circles*, CSLI Publications, Stanford.

BEALL, J.C., 1998, "Completing Sorensen's Menu: A Non-Modal Yabloesque Curry.", *Mind* 108: 737-39.

BEALL, J.C., 2001, "A Neglected Approach to the Liar", *Analysis* 61: 126-29.

BELNAP, N. D., 1982, "Gupta's Rule of Revision Theory of Truth", *Journal of Philosophical Logic* 11: 103-116.

BLACKBURN, S. & SIMMONS, K., (eds.), 1999, *Truth*, Oxford UP, Oxford.

BOOLOS, G. & JEFFREY, R., 1989, *Computability and Logic*, Cambridge UP, Cambridge, 3ª edición (1ª edición 1974).

BORGES, J. L., 1960, *El hacedor*, Emecé Editores, Buenos Aires.

BROMAND, J., 2002, "Why Paraconsistent Logic Can Only Tell Half the Truth", *Mind* 111: 741-49.

BURGE, T., 1979, "Semantical Paradox", *Journal of Philosophy* 76: 169-198. Reimpreso en MARTIN 1984, pp. 83-117 (referencias a esta reimpresión).

BURGE, T., 1982, "The Liar Paradox: Tangles and Chains". *Philosophical Studies* 41: 353-66.

CASABAN, E., (ed.), 2002, *Actes del XIVè Congrés Valencià de Filosofia (Peníscola, 21-23 marzo de 2002)*, Societat de Filosofia del País Valencià, València.

CHAPUIS, A., 2000. "Rationality and Circularity", en CHAPUIS & GUPTA 2000, pp. 49-78.

CHAPUIS, A. & GUPTA, A., 2000. *Circularity, Definition and Truth*, Indian Council of Philosophical Research (distribuido fuera de India por Ridgeview Publishing Company, Atascadero (California)), Nueva Delhi

CHIHARA, C. S., 1979, "The Semantic Paradoxes: a Diagnostic Investigation", *Philosophical Review* 88: 590-618.

CHIHARA, C. S., 1984a, "The Semantic Paradoxes: Some Second Thoughts", *Philosophical Studies* 45: 223-29.

CHIHARA, C. S., 1984b, "Priest, the Liar, and Gödel", *Journal of Philosophical Logic* 13: 117-24.

DEVLIN, K., 1993, *The Joy of Sets*, Springer Verlag, Nueva York, 2ª edición (1ª edición 1979).

DOWDEN, B. H., 1984, "Accepting Inconsistencies from the Paradoxes", *Journal of Philosophical Logic* 13: 125-30.

DUMMETT, M., 1978, *Truth and Other Enigmas*, Harvard UP, Cambridge (Massachusetts).

FALGUERA, J. L., RIVAS, U. & SAGÜILLO, J. M. (eds.), 1999, *Actas del congreso internacional: "La Filosofía Analítica en el Cambio de Milenio" (Santiago de Compostela, 1-4 diciembre 1999)*, Universidade de Santiago de Compostela publicacións, Santiago de Compostela.

FEFERMAN, S., 1984, "Towards Useful Type-Free Theories, I", *Journal of Symbolic Logic* 49: 75-111. Reimpreso en MARTIN 1984, pp. 237-87 (referencias a esta reimpresión).

FREGE, G., 1884, *Die Grundlagen der Arithmetik. Eine logisch mathematische Untersuchung über den Begriff der Zahl*. W. Koebner, Breslau. Traducción al castellano de Ulises Moulines: *Los fundamentos* *de la aritmética: investigación lógico-matemática sobre el concepto de número*, Laia, Barcelona, 1972.

FREGE, G., 1966, *Logische Untersuchungen*, Vandenhoeck und Ruprecht, Goettingen. Traducción al castellano de Luis Ml. Valdés Villanueva, *Investigaciones lógicas*, Tecnos, Madrid, 1984.

FREGE, G., 1918-19 "Der Gedanke. Eine logische Untersuchung". *Beiträge zur Philosophie des deutschen Idealismus* 1: 58-77. Traducción al castellano de Luis Ml. Valdés Villanueva: "El pensamiento: una investigación lógica", en *Investigaciones lógicas*, Tecnos, Madrid, 1984 (referencias a esta traducción).

GABBAY, D. M. & WANSING, H. (eds.), 1999, *What is Negation?*, Kluwer, Dordrecht.

GAIFMAN, H., 1992, "Pointers to Truth", *Journal of Philosophy* 89: 223-61.

GOLDSTEIN, L., 1994, "A Yabloesque Paradox in Set Theory", *Analysis* 54: 223-27.

GRATTAN-GUINESS, I., 1998, "Structural Similarities or Structuralism", *Mind* 107: 823-34.

GRICE, P., 1975, "Logic and Conversation", en P. Cole & J. L. Morgan (eds.) *Sintax and Semantics. Speech Acts*, vol 3, Academic Press, San Diego (California), 1975. Traducción al castellano de J. J. Acero: "Lógica y conversación" en VALDÉS VILLANUEVA 1991, pp. 511-30.

GRIM, P., 1998, Review of *Beyond the Limits of Thought*, by G. Priest, *Philosophy and Phenomenological Research* 58: 719-23.

GROVER, D. L., 1992, *A Prosentential Theory of Truth*, Princeton UP, Princeton.

GROVER, D. L., CAMP, J. L. & BELNAP, N., 1975, "A Prosentential Theory of Truth", *Philosophical Studies* 27: 73-125. Reimpreso en GROVER 1992, pp. 70-120.

GUPTA, A., 1982, "Truth and Paradox", *Journal of Philosophical Logic* 11: 1-60. Reimpreso en MARTIN 1984, pp. 175-235 (referencias a esta reimpresión).

GUPTA, A. & BELNAP, N., 1993, *The Revision Theory of Truth*, MIT Press, Cambridge (Massachusetts).

HAACK, S., 1978, *Philosophy of Logics*, Cambridge UP, Cambridge. Traducción al castellano de Amador Antón y Teresa Orduña: *Filosofía de las lógicas*, Cátedra, Madrid, 1982.

HALLETT, M., 1984, *Cantorian Set Theory and Limitations of Size*, Oxford UP, Oxford.

HALMOS, P. R., 1960, *Naive Set Theory*, D. Van Nostrand, Princeton (New Jersey). Traducción al castellano de Antonio Martín-Lunas: *Teoría intuitiva de los conjuntos*, Compañía Editorial Continental S.A., México D. F., 1965, (sexta impresión corregida, 1971, referencias a esta reimpresión de la traducción).

HARDY, J., 1995, "Is Yablo's Paradox Liar-like?", *Analysis* 55: 197-98.

HARDY, J., 1997, "Three Problems for the Singularity Theory of Truth", *Journal of Philosophical Logic* 26: 501-20.

HERZERBERGER, H. G., 1970a, "Paradoxes of Grounding in Semantics", *Journal of Philosophy* 67: 145-67.

HERZERBERGER, H. G., 1970b, "Truth and Modality in Semantically Closed Languages", en MARTIN 1978, pp. 25-46.

HERZERBERGER, H. G., 1982a, "Naive Semantics and the Liar Paradox.", *Journal of Philosophy* 79: 479-97.

HERZERBERGER, H. G., 1982b, "Notes on Naive Semantics", *Journal of Philosophical Logic* 11: 61-102. Reimpreso en MARTIN 1984, pp. 133-73 (referencias a esta impresión).

HORWICH, P., 1998, *Truth*, Oxford UP, Oxford, 2ª edición (1ª edición 1990).

KANT, I., 1978, *Crítica de la razón pura* (traducción de Pedro Ribas), Alfaguara, Madrid. (Originalmente publicado en alemán: *Kritik der reinen Vernunft*, 1781, 1ª edición.)

KEARNS, J. T., 1970, "Some Remarks Prompted by van Fraassen's Paper", en MARTIN 1978, pp. 47-58.

KIRKHAM, R. L., 1992, *Theories of Truth. A Critical Introduction*, MIT Press, Cambridge (Massachusetts).

KIRK, C. S., RAVEN, J. S. & SCHOFIELD, M., 1983, *The Presocratic Philosophers. A Critical History with a Selection of Texts*, Cambridge UP, Cambridge, 2ª edición (1ª edición 1957). Cito la traducción al castellano de Jesús García Fernández: *Los filósofos presocráticos. Historia crítica con selección de textos*, Gredos, Madrid, 1987, 2ª edición.

KLEENE, S. C., 1952, *Introduction to Metamathematics*, Van Nostrand, Nueva York. Traducción al castellano de M. Garrido y otros: *Introducción a la metamatemática*, Tecnos, Madrid, 1974.

KRIPKE, S., 1975, "Outline of a Theory of Truth", *Journal of Philosophy* 72: 690-716. Reimpreso en MARTIN 1984, pp. 53-82 (referencias a esta reimpresión).

KRIPKE, S., 1976, "A Theory of Truth I & II" (resúmenes), *Journal of Symbolic Logic* 41: 556-57.

LACKEY, D. (ed.), 1973, *Essays in Analysis*, Allen & Unwin, Londres.

LÖWE, B. et al. (eds.), 2003, *Foundations of the Formal Sciences II: Applications of Mathematical Logic to Philosophy and Linguistics*, Kluwer, Dordrecht.

MARTIN, R. L., 1970, "A Category Solution to the Liar", en MARTIN 1978, pp. 91-112.

MARTIN, R. L. (ed.), 1978, *The Paradox of the Liar*, Ridgeview, Reseda, 2ª edición (1ª edición 1970).

MARTIN, R. L. (ed.), 1984, *Recent Essays on Truth and the Liar Paradox*, Oxford UP, Oxford.

MARTIN, R. L. & WOODRUFF, P. W., 1975, "On Representing 'True-in-L' in L", *Philosophia* 5: 213-17. Reimpreso en MARTIN 1984, pp. 47-51 (referencias a esta reimpresión).

MARTÍNEZ FERNÁNDEZ, J., 1999, "Sobre la relación entre la paradoja de Russell y el mentiroso", en FALGUERA et al. 1999, pp. 87-94.

MARTÍNEZ FERNÁNDEZ, J., 2003, "The Gupta-Belnap Fixed-Point Problem and the Theory of Clones of Functions", en LÖWE, B. et al. 2003, pp. 175-84.

McGEE, V., 1989, "Applying Kripke's Theory of Truth", *Journal of Philosophy* 86: 530-9.

McGEE, V., 1991, *Truth, Vagueness and Paradox. An Essay on the Logic of Truth*, Hackett, Indianapolis.

McLARTY, C., 1993, "Anti-Foundation and Self-Reference", *Journal of Philosophical Logic* 22: 19-28.

PARSONS, C., 1974, "The Liar Paradox", *Journal of Philosophical Logic* 3: 381-412. Reimpreso en MARTIN 1984, pp. 9-45 (referencias a esta reimpresión).

PEANO, G., 1906, "Addition E", *Rivista di Mathematica* 8: 143-57.

PRIEST, G., 1979, "The Logic of Paradox", *Journal of Philosophical Logic* 8: 219-41.

PRIEST, G., 1984a, "Logic of Paradox Revisited", *Journal of Philosophical Logic* 12: 153-79.

PRIEST, G., 1984b, "Hyper-Contradictions", *Logique et Analyse* 27: 237-43.

PRIEST, G., 1987, *In Contradiction. A Study of the Transconsistent*, M. Nijhoff, Dordrecht.

PRIEST, G., 1991, "Minimally Inconsistent LP", *Studia Logica* 50: 321-31.

PRIEST, G., 1993, "Can Contradictions be True? - II", *Proceedings of the Aristotelian Society, suppl. vol.* 67: 35-54.

PRIEST, G., 1994a, "The Structure of the Paradoxes of Self-Reference", *Mind* 103: 25-34.

PRIEST, G., 1994b, "Is Arithmetic Consistent?", *Mind* 103: 337-49.

PRIEST, G., 1997a, "Yablo's Paradox", *Analysis* 57: 236-42.

PRIEST, G., 1997b, "Inconsistent Models of Arithmetic Part I: Finite Models", *Journal of Philosophical Logic* 26: 223-35.

PRIEST, G., 1998a, "To Be *and* Not to Be –That Is the Answer: On Aristotle on the Law of Non Contradiction", *Philosophiegeschichte und Logische Analyse* I: 91-130.

PRIEST, G., 1998b, "What is so Bad about Contradictions?", *Journal of Philosophy* 95: 410-26.

PRIEST, G., 1998c, "The Import of Inclosure: some Comments on Grattan-Guinness", *Mind* 107: 835-40.

PRIEST, G., 1999, "What Not? A Defence of a Dialetheic Account of Negation", en GABBAY & WANSING 1999, pp. 101-20.

PRIEST, G., 2000a, "Truth and Contradiction", *Philosophycal Quarterly* 50: 305-19.

PRIEST, G., 2000b, "Could Everything Be True?", *Australasian Journal of Philosophy* 78: 189-95.

PRIEST, G., 2000c, "Inconsistent Models of Arithmetic Part II: The General Case" *Journal of Symbolic Logic* 65: 1519-29.

PRIEST, G., 2000d, "On the Principle of Uniform Solution: A Reply to Smith", *Mind* 109: 123-5.

PRIEST, G., 2002, *Beyond the Limits of Thought*, Oxford UP, Oxford, 2ª edición (1ª edición 1995, Cambridge UP, Cambridge).

PRIEST, G., 2003, "On Alternative Geometries, Arithmetics, and Logics; a Tribute to Lukasiewicz", *Studia Logica* 74: 441-68.

QUINE, W. V. O., 1970, *Philosophy of Logic*, Prentice Hall, Englewood Cliffs. Traducción al castellano de Manuel Sacristán: *Filosofía de la lógica*, Alianza Editorial, Madrid, 1973.

QUINE, W. V. O., 1976a, *The Ways of Paradox and Other Essays*, Harvard UP, Cambridge (Massachusetts), 2ª edición revisada (1ª edición 1966).

QUINE, W. V. O., 1976b, "The Ways of Paradox", en Quine 1976a, pp. 1-18 (publicado por primera vez en 1962 bajo el título "Paradox" en *Scientific American* 206). Traducción al castellano de M. de Guzmán: "Paradoja" en *Matemáticas en el mundo moderno*, Blume, Madrid, 1974, pp. 224-33.

QUINE, W. V. O., 1980a, *From a Logical Point of View*, Harvard UP, Cambridge (Massachusetts), 2ª edición revisada (1ª edición 1953). Traducido al castellano por Manuel Sacristán: *Desde un punto de vista lógico*, Ariel, Barcelona, 1962 (reimpreso en Orbis, Barcelona, 1984).

QUINE, W. V. O., 1980b, "Two dogmas of empiricism" en Quine 1980a, pp. 20-46 (publicado por primera vez en *Philosophical Review*, 1951).

RAMSEY, F. P., 1925, 'The Foundations of Mathematics', *Proceedings of the London Mathematics Society* 25: 338-84. Reimpreso en RAMSEY 1931, pp. 1-61 (referencias a esta reimpresión).

RAMSEY, F. P., 1927, "On facts and Propositions", *Proceedings of the Aristotelian Society, suppl. vol.* 7: 153-70. Reimpreso en RAMSEY 1931, pp. 138-55.

RAMSEY, F. P., 1931, *The Foundations of Mathematics and Other Logical Essays*, (ed. R. B. Braithwaite) Routledge & Kegan Paul, Londres.

RUSSELL, B., 1905a, "On Denoting", *Mind* (New Series) 14: 479-93. Reimpreso en RUSSELL 1956, pp. 41-56 (referencias a esta reimpresión).

RUSSELL, B., 1905b, "On Some Difficulties in the Theory of Transfinite Numbers and Order Types", *Proceedings of the London Mathematical Society* (series 2) 4: 29-53. Reimpreso en LACKEY 1973 (referencias a esta reimpresión).

RUSSELL, B., 1908, "Mathematical Logic as based on the Theory of Types", *American Journal of Mathematics* 30: 222-62. Reimpreso en RUSSELL 1956, pp. 59-102 (referencias a esta reimpresión).

RUSSELL, B., 1956, *Logic and Knowledge* (ed. R. C. Marsh), Allen & Unwin, Londres. Reimpreso en 1992 por Routledge, Londres. Traducción al castellano de Javier Muguerza: *Lógica y conocimiento*, Taurus, Madrid, 1981.

SAGÜILLO, J. M., FALGUERA, J. L. & MARTÍNEZ, C. (eds.), 2001, *Actas del congreso internacional: "Teorías formales y teorías empíricas: aspectos fundacionales" (Santiago de Compostela, 14-16 de noviembre de 2001)*, Universidade de Santiago de Compostela publicacións, Santiago de Compostela.

SAINSBURY, R. M., 1995, *Paradoxes*, Cambridge UP, Cambridge, 2ª edición (1ª edición 1987).

SKYRMS, B., 1970a, "Return of the Liar: Three-Valued Logic and the Concept of Truth", *American Philosophical Quarterly* 7: 153-61.

SKYRMS, B., 1970b, "Notes on Quantification and Self-Reference", en MARTIN 1978, pp. 67-74.

SHAPIRO, S., 2002, "Incompletness and Inconsistency", *Mind* 111: 817-32.

SIMMONS, K., 1993, *Universality and the Liar. An Essay on Truth and the Diagonal Argument*, Cambridge UP, Cambridge.

SMILEY, T., 1993, "Can Contradictions Be True? - I", *Proceedings of the Aristotelian Society, suppl. vol.* 67: 17-33.

SMITH, N. J. J., 2000, "The Principle of Uniform Solution (of the Paradoxes of Self-Reference)", *Mind* 109: 117-22.

SMULLYAN, R. M., 1987, "Quotation and Self-Reference", en BARTLETT & SUBER 1987, pp. 123-44.

SMULLYAN, R. M., 1992, *Gödel's Incompleteness Theorems*, Oxford UP, Oxford.

SMULLYAN, R. M., 1994, *Diagonalization and Self-Reference*, Oxford UP, Oxford.

SOAMES, S., 1999, *Understanding Truth*, Oxford UP, Oxford.

SORENSEN, R. A., 1998, "Yablo's Paradox and Kindred Infinite Liars", *Mind* 107: 135-55.

STERN, R. (ed.), 1999, *Transcendental Arguments: Problems and Prospects*. Oxford UP, Oxford.

STRAWSON, P. F., 1950, "Truth". *Proceedings of the Aristotelian Society, suppl. vol.* 24: 129-56. Traducción al castellano de Alfonso García Suárez y Luis Ml. Valdés Villanueva: "Verdad" (en P. F. STRAWSON, *Ensayos lógico-lingüísticos*, Tecnos, Madrid, 1983.)

TAPPENDEN, J., 1993, "The Liar and Sorites Paradoxes: Toward a Unified Account", *Journal of Philosophy* 90: 551-77.

TARSKI, A., 1933, *Projecie prawdy w jezykach nauk dedukcyjnych* (Sobre el concepto de verdad en los lenguajes de las ciencias deductivas), Varsovia. Traducción al alemán "Der Wahrheitsbegriff in den formalisierten Sprachen", *Studia Philosophica* 1: 261-405, 1935. Traducción al inglés: "The Concept of Truth in Formalized Languages", en TARSKI 1983, pp. 152-278 (referencias a esta traducción).

TARSKI, A., 1944, "The Semantic Conception of Truth and the Foundations of Semantics", *Philosophy and Phenomenological Research* 4: 341-375. Traducción al castellano de M. Bunge et al.: "La concepción semántica de la verdad" en M. BUNGE (ed.), *Antología semántica*, Nueva Visión, Buenos Aires, 1960. Reimpreso en VALDÉS VILLANUEVA 1991, pp. 275-313 (referencias a esta reimpresión).

TARSKI, A., 1983, *Logic, Semantics and Metamathematics* (ed.: J. Corcoran), traducción al ingles de J. H. Woodger, Hackett, Indianapolis, 2ª edición (1ª edición, 1956).

TENNANT, N., 1995, "On paradox without self-reference", *Analysis* 55: 199-207.

TENNANT, N., 1998, "Critical Notice of *Beyond the Limits of Thought*", *Philosophical Books* 39: 20-38.

TENNANT, N., 1999, "Negation, Absurdity and Contrariety", en GABBAY & WANSING 1999, pp. 199-222.

VALDÉS VILLANUEVA, L. ML., 1991, *La búsqueda del significado. Lecturas de filosofía del lenguaje*, Tecnos-Universidad de Murcia, Madrid.

VALOR ABAD, J., 2001, "Grounded Sentences", en SAGÜILLO et al. 2001, pp. 531-43.

VALOR ABAD, J., 2002, "Veritat y circularitat", en CASABAN 2002, pp. 123-39.

VAN FRAASSEN, B. C., 1968, "Presupposition, Implication and Self-Reference". *Journal of Philosophy* 65: 136-52.

VAN FRAASSEN, B. C., 1970a, "Truth and Paradoxical Consequences", en MARTIN 1978, pp. 13-23.

VAN FRAASSEN, B. C., 1970b, "Rejoinder: On a Kantian Conception of Language", en MARTIN 1978, pp. 59-66.

VISSER, A., 1984, "Four Valued Semantics and the Liar", *Journal of Philosophical Logic* 13: 181-212.

VISSER, A., 1989, "Semantics and the Liar Paradox", en *Handbook of Philosophical Logic vol. 4* (ed. D. M. Gabbay), D. Reidel, Dordrecht, pp. 617-706.

WILLIAMSON, T., 1996, Review of *Beyond the Limits of Thought*, by G. Priest, *British Journal for the Philosophy of Science* 47: 331-34.

WITTGENSTEIN, L., 1921, *Tractatus Logico-Philosophicus*, en *Annalen*, cuaderno 14 (y edición bilingüe (alemán-inglés) con traducción al inglés de C. K. Ogden y F. P. Ramsey en Kegan Paul, Londres, 1922). Traducción al castellano de Jacobo Muñoz e Isidoro Reguera, Alianza Editorial, Madrid, 1973 (referencias a esta traducción).

WITTGENSTEIN, L., 1953, *Philosophische Untersuchungen-Philosophical Investigations* (edición bilingüe (alemán-inglés) con traducción al inglés de G. E. M. Anscombe), Blackwell, Oxford. Traducción al castellano de Alfonso García Suárez y Ulises Moulines: *Investigaciones filosóficas*, Instituto de Investigaciones Filosóficas de la UNAM, México D. F./Crítica, Barcelona, 1988 (referencias a esta traducción).

WEIR, A., 1999, Review of *Beyond the Limits of Thought*, by G. Priest, *Philosophical Quarterly* 49: 122-25.

YABLO, S., 1982, "Grounding Dependence and Paradox", *Journal of Philosophical Logic* 11: 117-37.

YABLO, S., 1985, "Truth and Reflection", *Journal of Philosophical Logic* 14: 297-349.

YABLO, S., 1989, "Truth, Definite Truth, and Paradox", *Journal of Philosophy* 86: 539-41.

YABLO, S., 1993, "Paradox without Self-Reference", *Analysis* 53: 251-52.

YACUB, A., 1993, *The Liar Speaks the Truth*, Oxford UP, Oxford.

ZADEH, L. A., 1975, "Fuzzy Logic and Approximate Reasoning", *Synthese* 32: 407-28.