Paradojas del Mentiroso y Dialeteísmo

Un análisis del tratamiento de la paradoja del mentiroso en la motivación del dialeteísmo de Graham Priest

Pablo Rivas-Robledo

Tesis de grado presentada para obtener el título de
Filósofo

Director: Juan Camilo Espejo-Serna
Miembros del Comité de Evaluación: John Anderson Pinzón Duarte
Tomás Andrés Barrero Guzmán

Facultad de Filosofía y Ciencias Humanas
Universidad de La Sabana
Chía, Cundinamarca, Colombia
2020

Agradecimientos

Hay muchas personas que de una manera u otra contribuyeron al desarrollo de este trabajo con las cuales estoy agradecido. Quiero hacer unas menciones especiales.

Por encima de todo, gracias a Juan Camilo Espejo-Serna por enseñarme tantas, tantas cosas y por aventurarse en el mar de las contradicciones. Igualmente gracias a los dos ángeles, Miguel Ángel Quintero y Miguel Ángel Prieto por estar durante todo este tiempo viviendo la construcción de este trabajo tras bambalinas. Migue Prieto tuvo que aguantarse un gran número de memes y chistes a horas poco convenientes de la madrugada que yo creía que venían al caso, pero él se encargó de hacerme entrar en razón.

Gracias a Deathspell Omega, Emperor y Misþyrming por horas de buena música y gracias a la comunidad filosófica de la Universidad de La Sabana (el Ágora). Especialmente a Katherin Jiménez y a Martín Buenahora. En Martín encontré un pasión común por Leibniz y LaTeX(que muchas veces fue de gran utilidad, pero muchas otras fueron fuente de discusión) y con quien pude compartir muchos chistes malos, los cuales reafirmaron mi creencia de que la mejor manera de llegar a la verdad es riendo.

También quiero agradecer a las personas que leyeron partes de este trabajo, especialmente a Juliana Ocampo. De igual manera quiero agradecer a Richard Kimberly Heck y a Lavinia Picollo, quienes contestaron mis preguntas sin estar obligados a hacerlo. Un agradecimiento también a mis amigos de PostMortem FetalExtrusion y, sin duda, un agradecimiento especial a Alexandra Elbalkyan. Por último quiero agradecer a Fabio Enrique Pulido-Ortiz, quien ha sido un gran mentor, y a los miembros del jurado por permitirme mostrarles algo de mi labor filosófica, especialmente a Anderson Pinzón, quien ha sido testigo de mi paso por la carrera desde primer semestre. Espero haber aprendido una o dos cosas.

Índice general

Capítulo 1

Introducción

Con el advenimiento de la lógica moderna a principios del siglo XX, Frege y Whitehead-Russell incorporaron en sus lógicas el principio de explosión o Ex Quodlibet Contradictione (ECQ), según el cual de dos afirmaciones contradictorias entre sí se sigue cualquier otra afirmación.[1] Presumiblemente, este principio tiene su origen en el siglo XII en la lógica de William de Soissons (Martin, 1986, p. 565) y tuvo una gran acogida entre pensadores posteriores, como Alexander de Neckham y el Pseudo-Escoto (Priest, 2007, p. 133). Como muchos temas de la lógica medieval, el principio de explosión pasó desapercibido en occidente hasta su reaparición en el siglo XX.

Los sistemas lógicos que aceptan el principio de explosión han dominado la historia de la lógica en occidente, pues este fue incorporado a numerosos sistemas posteriores al de los *Principia Mathematica* y a partir de la publicación de *Symbolic Logic* por C.I. Lewis y C.H. Langford el principio de explosión se utilizó para rechazar varios sistemas lógicos en tanto estos llevaban a contradicciones y por lo tanto en estos se podía probar cualquier cosa (Bobenrieth Miserda, 2010, p. 112). Este paradigma empezó a cambiar con el advenimiento de las lógicas paraconsistentes, sistemas lógicos en los cuales se admitían

[1] En la lógica de Frege, el principio se presenta como la proposición 36 del *Begriffsschrift*. En la lógica de Whitehead-Russell el principio se presenta como la consecuencia 2.21 del *Principia Mathematica*. En Bobenrieth Miserda (2010; 1996, cap. IV) se encuentra una explicación más detallada del lugar que tiene el principio de explosión en los *Principia*, además de un análisis del principio en los sistemas de Post, Hilbert & Ackermann y Łukasiwiecz, que son contemporáneos a los de Frege y Whitehead-Russell.

algunas contradicciones y a partir de estas se podían hacer inferencias no triviales a partir de estas. Más allá de las discusiones históricas sobre los antecedentes y pre-historia de la lógica paraconsistente, hoy se considera que el primer sistema de lógica paraconsistente es el sistema C_1 introducido por Newton da Costa en (1963) (ver da Costa (1974) para una versión en inglés de algunas de las ideas expresadas allí.[2]

Los trabajos de da Costa se vieron nutridos con la participación de varios colegas a lo largo de los años (Bobenrieth Miserda, 1996, cap. X; Priest, 2007, sec. 4.3), pero rara vez estos trabajos iban más allá de las consideraciones matemáticas y lógicas, dejando de lado las implicaciones filosóficas. De hecho, el mismo trabajo de da Costa se desarrolló porque los sistemas paraconsistentes simplemente le parecían interesantes y los consideraba teorías que posiblemente eran importantes para las matemáticas (Priest, 2007, p. 167).

Paralelo al desarrollo de la lógica paraconsistente, se va gestando una corriente filosófica que en quería defender la existencia de contradicciones y explorar las consecuencias filosóficas de las lógicas paraconsistentes. Es así como a mediados de los años 70 surge el dialeteísmo. El dialeteísmo defiende que existen *dialetheiai* (plural de dialetheia), enunciados que son verdaderos pero su negación también es verdadera. [3]

Esto significa (1) que existen contradicciones que son verdaderas y (2) que el Ex Quodlibet Contradictione o ley de la explosión (según la cual de una contradicción se sigue cualquier cosa) no es una ley lógica, pues tiene contraejemplos: las dialetheiai. Si bien los antecedentes del dialeteísmo pueden rastrearse hasta la filosofía de G.W.F. Hegel, los primeros en desarrollar sus postulados principales fueron Richard Routley y Graham Priest.

Routley, quien luego tomaría el apellido Sylvan, murió en 1996 dejando un ambicioso proyecto filosófico con el dialeteísmo como uno de sus componentes principales. Priest, por su lado, continuó desarrollando y defendiendo el dialeteísmo de sus críticos, que no han sido pocos. Es así como el dialeteísmo se ha consolidado como una manera particular de

---

2 La lógica trivalente de Łukasiewicz, la lógica imaginaria de Vasiliev y la lógica discursiva de Jaśkowski son normalmente citadas como precursoras de la lógica paraconsistente. Sin embargo, estas lógicas siguen siendo explosivas de una forma u otra (tal vez salvo con la excepción de la lógica discursiva de Jaśkowski) (Priest, 2007, pp. 146-147, sec. 4.2).

3 El término *dialeteísmo* apareció por primera vez en Priest & Routley (1989).

tomar las contradicciones: en vez de rechazarlas o resolverlas, como ha hecho gran parte de la filosofía en occidente, el dialeteísmo sostiene que algunas de estas son verdaderas contradicciones, es decir enunciados verdaderos, de los cuales no se sigue cualquier cosa.

A lo largo de los años, Priest ha hecho aportes en varios campos de la filosofía, primordialmente en lógica y metafísica, pero también en otras áreas como la epistemología, la filosofía del lenguaje, la mereología, la historia de la filosofía y la filosofía de las matemáticas.[4] A pesar de la diversidad de temas, el dialeteísmo es el hilo conductor de toda la obra de Priest, pues la inmensa mayoría de su trabajo filosófico es o bien una defensa de este o una aplicación de sus postulados.

Para Priest, la principal motivación para el dialeteísmo es precisamente cómo este puede dar una solución uniforme a una serie de paradojas denominadas paradojas de la autorreferencia. En *Beyond the Limits of Thought*, Priest (1995/2002, caps. 9-10) ofrece un minucioso análisis de las paradojas de la autorreferencia, como la paradoja de Russell-Zermelo o la paradoja del mentiroso, y argumenta que todas estas comparten la misma estructura (lo que él llama *Inclosure Schema*), por lo que deberían compartir el mismo tipo de solución. Esto último se conoce como el principio de solución uniforme (PUS, por sus siglas en inglés) (Priest, 1995/2002, secs. 11.5-11.6). Según Priest, la única aproximación uniforme y satisfactoria a la luz de la estructura de las paradojas y el PUS es una aproximación dialeteísta: tomar estas paradojas como contradicciones verdaderas (Priest, 1995/2002, p. 169). Las paradojas se resuelven conservando la contradicción.

De esta manera, si bien el desarrollo filosófico de Priest es amplio, todo esto fue motivado por su interpretación específica de las paradojas de la autorreferencia en *Beyond the Limits of Thought*. Esto es especialmente cierto no solo para su versión del dialeteísmo, sino el dialeteísmo como una corriente filosófica, pues la aproximación de las paradojas de la autorreferencia hecha por Priest en *Beyond the Limits of Thought* es considerada como el argumento maestro del dialeteísmo para demostrar la existencia de dialetheiai y motivar el dialeteísmo de manera satisfactoria (Priest y col., 2018, sec. 3.1).

---

[4] Más sobre esto en el capítulo 2.

1.1 La paradoja del mentiroso en la motivación de Priest

Puede que todas las paradojas de la autorreferencia tengan la misma estructura y por lo tanto la misma solución, como afirma Priest. Pero en esta tesis ofrezco razones para pensar que Priest no ha demostrado que todas las paradojas de la autorreferencia tienen la misma estructura, es decir, que encajen en el *Inclosure Schema*. Y esto no es porque le falte analizar algunas paradojas, sino porque comete un error al analizarlas.

En concreto, Priest sostiene que todas las distintas formulaciones de la paradoja del mentiroso son exactamente la misma paradoja, es decir, no solo son el mismo tipo de paradoja (una paradoja de autorreferencia, como la de Russell-Zermelo), sino que son exactamente la misma paradoja y que tan solo con demostrar que una de las formulaciones tiene la misma estructura que otras paradojas de la autorreferencia, ya demuestra que todas las formulaciones de la paradoja del mentiroso tienen la misma estructura. En su tratamiento de la paradoja del mentiroso en *Beyond the Limits of Thought*, Priest analiza principalmente tres formulaciones de la paradoja del mentiroso: $\sigma = \langle \sigma \notin Tr \rangle$, 'estoy mintiendo' y 'esta oración es falsa'. Como hago evidente en el capítulo 2, Priest solo demuestra cómo la primera versión encaja dentro del *Inclosure Schema* y va de una a otra versión de la paradoja indistintamente.

En concreto, Priest se abstiene de demostrar cómo las otras formulaciones encajan en el *Inclosure Schema* porque piensa que las diferencias entre las formulaciones son accidentales y al fin de cuentas todas estas son la misma paradoja:[5]

> The Liar paradox comes in many varieties. In some of them, self-reference is achieved by means of a demonstrative ('this sentence', etc.); in others is obtained by a description ('the first sentence on such and such page', etc.); yet in others is obtained by some diagonal argument using Gödel-numbers as names. Now, at some level of abstraction, these paradoxes are, presumably, of different kinds. The first, after all, depends on context-dependent features of utterance; the second presupposes some account of the way that descrip-

[5] Por motivos de claridad prefiero poner las citas directas en inglés.

> tions work; the third employs numbers. (...) **The exact *mechanisms* of reference are, in a clear sense, accidental to what is going on.** (...) In all versions of the Liar paradox, there is a sentence, α, such that: α iff $\langle\alpha\rangle$ is false (or not true) (Priest, 1995/2002, p. 290, la cursiva es del original, la negrilla es mía).

Esta forma de tratar las paradojas me parece injustificada. Considero que las diferencias entre las formulaciones de la paradoja no son accidentales, por lo que no puede decirse que habiendo demostrado que una formulación encaja en el esquema ya se han encajado todas. En este trabajo pretendo dar buenas razones para esto.

Estos *saltos* entre paradojas, es decir, usarlas indistintamente es una costumbre arraigada en la filosofía occidental, por ejemplo, Barwise y Etchmendy van de 'estoy mintiendo' a 'esta oración es falsa' sin detenerse a justificar este salto (1987, p. 3); Beall, Glanzberg y Ripley van de una paradoja a otra, examinando estas como miembras de una misma familia, pero no establecen qué hace a una formulación de la paradoja distinta a otra ni por qué deben considerarse como un una familia (2019, Introduction, secs. 1-2). Sin embargo, en Priest estos saltos son especialmente problemáticos porque implican que a consciencia está ignorando diferencias entre las paradojas que, a pesar de juzgar como accidentales, son esenciales.

1.2 Qué quiero hacer en este trabajo y cómo planeo hacerlo

En este trabajo quiero dar buenas razones a favor de la siguiente afirmación: las diferencias entre las distintas formulaciones de la paradoja no son accidentales y por lo tanto no son la misma paradoja en el sentido en el Priest sostiene que lo son. Para ello requiero (1) exponer el tratamiento que le da Priest a la paradoja del mentiroso y a otras paradojas de la autorreferencia en *Beyond the Limits of Thought*; y (2) demostrar que hay diferencias entre las formulaciones de la paradoja que las hacen irreducibles entre sí.

Para ello, dedico el segundo capítulo a explicar con detalle el *Inclosure Schema* y cómo una serie de paradojas de la autorreferencia encajan en este, esto lo hago siguiendo atentamente los capítulos 8 a 11 y 17 de *Beyond the Limits of Thought* y una serie de comentarios que Priest hizo sobre el tema en trabajos posteriores. Un punto importante del capítulo 2 es identificar las formulaciones de la paradoja que Priest está teniendo en cuenta, una en lenguaje formal ($\sigma = \langle \sigma \notin Tr \rangle$) y otras dos en lenguaje natural ('estoy mintiendo' y 'esta oración es falsa').

El tercer capítulo se encarga de exponer de manera precisa cómo es posible formular la paradoja en lenguaje formal al estilo de Priest y presento algunos apuntes sobre el concepto de referencia en lenguajes formales. Así, este capítulo contiene algunas consideraciones necesarias sobre filosofía de la lógica y de las matemáticas con las cuales pretendo acercarme a fórmulas lógicas a través de una teoría del significado. Esta es una aproximación que es apenas naciente y se nutre de los trabajos de Lavinia Picollo, Hannes Leitgeb y Richard Kimberly Heck. Este capítulo es probablemente el más complejo, dado que requiere de un aparato formal de una alta complejidad, el cual me esfuerzo por hacer lo más claro posible.

Una vez explorada la paradoja del mentiroso en lenguaje formal, en el cuarto capítulo paso a explicar el mecanismo mediante el cual es posible formular las paradojas en lenguaje natural ('estoy mintiendo' y 'esta oración es falsa'). Dado que en este capítulo se analizan dos paradojas es el más extenso de todo el trabajo. Para poderme aproximar a estas formulaciones de la paradoja hago uso extensivo de la teoría de los deíciticos de David Kaplan y la teoría de las intenciones directrices de John Perry, las cuales cobijo bajo un mismo término como semántica de Kaplan-Perry.

Los resultados de los capítulos precedentes permiten inferir lo siguiente: Priest pretende reducir las formulaciones de la paraodoja del mentiroso a solo una, la del lenguaje formal. Esta formulación hace uso de un mecanismo de autorreferencia conocido como diagonalización o construcción diagonal que en este caso se obtiene mediante a partir de una fórmula cuyo cuantificador principal es existencial afirmativo la cual crea una paradoja que hace referencia a fórmulas del lenguaje formal mediante números de Gödel. La primera formulación de la paradoja en lenguaje natural ('estoy mintiendo') hace uso del indéxico 'yo' para lograr autorreferencia y refiere a oraciones dichas en un contexto por un agente particular. La segunda formulación de la paradoja en lenguaje natural ('esta oración es falsa') hace uso del demostrativo 'esta' para lograr autorreferencia y refiere a oraciones del lenguaje natural.

En relación con estos hallazgos, en el quinto capítulo propongo que el análisis de las paradojas del mentiroso proporcionado por Priest no es del todo correcto, lo cual crea un problema para su versión del dialeteísmo. Según Priest, las diferencias entre las versiones de la paradoja del mentiroso son accidentales, por lo que es posible reducir las otras formulaciones de la paradoja en lenguaje natural a la formulación en lenguaje formal. Como demuestro en el quinto capítulo, al tratarlas como accidentales, Priest pierde la motivación maestra para el dialeteísmo, según la cual las paradojas deben tener una solución uniforme, pues solo haría falta una solución y no un tipo de solución.

Es decir, Priest nos obliga a ignorar diferencias esenciales entre tres paradojas del mentiroso distintas y a tomarlas como la misma paradoja. Si esto es así, entonces esto mismo se podría decir de otras paradojas que cumplen con el PUS: todas cuentan con diferencias accidentales. Estas paradojas dejarían de ser el mismo tipo de paradoja y serían la misma paradoja. De este modo, todas las paradojas de la autorreferencia serían la misma, por lo tanto no tendría necesidad de formular ni el *Inclosure Schema* ni el PUS y todas las paradojas tendrían exactamente la misma solución, e.g. la teoría de conjuntos de Zermelo-Fraenkel, y se hace innecesaria una aproximación dialeteísta.

Pienso que mientras se hagan algunas concesiones, esta objeción no es insuperable, ni tiene el propósito de serlo. De esta manera, termino este trabajo ofreciendo algunas conclusiones, donde explico cuáles son las concesiones que considero que el dialeteísmo de Priest tiene que hacer para superar esta objeción y planteo algunas tareas pendientes para esta corriente. Así mismo, en la última sección ofrezco algunas implicaciones que tiene este trabajo más allá del dialeteísmo de Priest y comento acerca de futuras investigaciones que pueden nacer a partir de lo hecho en este trabajo.

Capítulo 2

Dialeteísmo, el *Inclosure Schema*, el Principio de Solución Uniforme y la paradoja del mentiroso

"*The Gödel and liar sentences shouldn't be compared with a person saying: 'I lost the script'; it should be compared with somebody saying: 'The last person lost the script.' If we describe the self-referentiality of the liar sentence by seeing it as a sentence that says about itself that it is not true, then we ascribe a kind of self-referentiality to it that it doesn't have.*"

Sophia, en el diálogo *The Root of Evil*

(Halbach, 2016, p. 162)

A lo largo de la historia de la filosofía han aparecido numerosas paradojas, pero hacia principios del siglo XX, con la aparición de lenguajes formales más poderosos, hubo un renovado interés en el análisis filosófico y matemático de las mismas (Cantini, 2009). A lo largo de este siglo aparecieron numerosos de sistemas formales y herramientas filosóficas que se enfrentaron a las paradojas. El dialeteísmo surge como una respuesta a algunos de estos sistemas.

Previa a la aparición del dialeteísmo, dos planteamientos respecto a las paradojas venían siendo aceptados sin mayores complicaciones. El primero era una cierta aversión por las paradojas, se les tomaba como algo indeseable o como evidencia de error en una teoría o en un argumento (Priest, 1995/2002, p. 4). Esta postura ha permeado toda la filosofía occidental desde Aristóteles (Priest y col., 2018, sec. 2.1; Priest, 2007, p. 139). El segundo es una tendencia a clasificar las paradojas entre paradojas lógicas o matemáticas y paradojas sobre conceptos que no son ni matemáticos ni lógicos. Esta forma de clasificar las paradojas empezó con Ramsey (1925) y se mantiene hasta nuestros días (ver, entre otros Bolander (2017)).

El dialeteísmo de Priest y Routley se opone a estas dos afirmaciones, y por eso se considera que choca con la ortodoxia filosófica en occidente. Es por eso que gran parte del trabajo de los dialeteístas es la fundamentación de su teoría. Priest se ha preocupado especialmente por fundamentar de la mejor manera posible el dialeteísmo, buscando una lógica apropiada para el mismo y la mejor forma de motivar su ruptura con la ortodoxia.

Temáticamente, la obra de Priest podría organizarse de la siguiente manera. En *Beyond the Limits of Thought*, Priest (1995/2002) expone su motivación para el dialeteismo: la aversión por las paradojas (especialmente por las de la autorreferencia) desde Aristóteles es completamente injustificada y el tratamiento separado que se les ha dado está condenado a fracasar, las paradojas de la autorreferencia son verdaderas contradicciones que tienen la misma estructura y que por lo tanto deberían tener el mismo tipo de solución. En *Beyond the Limits of Thought*, Priest sugiere que el dialeteísmo ofrece una solución uniforme a todas estas paradojas: considerarlas dialetheiai, enunciados que son tanto verdaderos como falsos. Sin embargo, allí Priest no ofrece una defensa ni un desarrollo de esta posición, pues su propósito es simplemente el de motivar el dialeteísmo a partir de la necesidad de una solución uniforme para las paradojas de la autorreferencia.

La defensa que Priest ofrece del dialeteísmo se encuentra en otros dos libros, *In Contradiction* (Priest, 1987/2006) y *Doubt Truth to be a Liar* (Priest, 2006). En el primero de estos, Priest presenta su defensa más completa del dialeteísmo hasta el momento, formulando un sistema lógico para las dialetheiai (el sistema LP) y discute algunos aspectos de su epistemología y metafísica. El segundo de estos textos continúa la defensa del dialeteísmo ahondando en las implicaciones que el dialeteísmo tiene para la naturaleza de la lógica y conceptos como la verdad, la racionalidad y la negación.

A partir de su dialeteísmo Priest pasaría a tratar una serie de problemas clásicos de la filosofía occidental y oriental en trabajos posteriores. En *One: Being an Investigation into* *the Unity of Reality and of its Parts, including the Singular Object which is Nothingness*, Priest (2014) explora varios problemas clásicos de la filosofía occidental, primordialmente el de los universales, mientras que en *Towards Non-Being* (Priest, 2005/2016) desarrolla una teoría de la intencionalidad basado en su trabajo y el de Richad Routley (Sylvan) sobre lógicas paraconsistentes y el dialeteísmo. Por último, en 2018 Priest publicó *The Fifth Corner of Four: an Essay on Buddhist Metaphysics and the Catuṣkoṭi* (Priest, 2018), un exploración de la metafísica budista bajo la perspectiva del dialeteísmo y la lógica paraconsistente, los cuales son fundamentales para explicar el catuṣkoṭi, un principio metafísico según el cual todas las aserciones son verdaderas, falsas, ambas o ninguna.

Como es evidente, las motivaciones para el dialeteísmo son especialmente relevantes para todo el sistema. Es por eso que en este capítulo (1) explico la motivación por excelencia del dialeteísmo de Graham Priest, la estructura uniforme de las paradojas de la autorreferencia y el principio de solución uniforme y (2) expongo el papel que juega la paradoja del mentiroso dentro de la motivación del dialeteísmo de Priest. Con esto, pretendo señalar el papel especial que juega la paradoja del mentiroso en el dialeteísmo de Priest

2.1 Las paradojas de la autorreferencia y el *Inclosure Schema*

La relación entre las paradojas de la autorreferencia y la motivación del dialeteísmo de Priest puede resumirse de la siguiente manera: las paradojas de la autorreferencia tienen la misma estructura, por lo tanto deberían tener el mismo tipo de solución, la única aproximación uniforme y satisfactoria a la solución de estas paradojas la ofrece el dialeteísmo: tomar las afirmaciones paradójicas como contradicciones verdaderas (Priest, 1995/2002, p. 169).

Para probar que las paradojas de la autorreferencia tienen la misma estructura, Priest recoge un número importante de estas (inicialmente 10) y construye un esquema en el que todas encajen, llamado *Inclosure Schema*.[6] De esta manera, si un conjunto de afirmaciones

[6] Priest inspira su tratamiento de las paradojas de la autorreferencia en varios comentarios dichos por

autorreferenciales contradictorias encajan en el esquema y estas no son por lo menos *prima facie* falsas, entonces se trata de una paradoja (Priest, 1995/2002, secs. 11.5-11.6, 17.2).

El *Inclosure Schema* consta de dos afirmaciones: en la primera se establece la existencia de un conjunto Ω conformado por los elementos del dominio que satisfagan la propiedad ϕ y que el conjunto satisface la propiedad ψ tal que $\psi(\Omega)$; en la segunda afirmación se establecen dos condiciones: si x es un subconjunto de Ω tal que $\psi(x)$, entonces (1) cualquier operación sobre el subconjunto no pertenece al subconjunto ($\delta(x) \notin x$), pero (2) cualquier operación sobre el subconjunto sí pertenece al conjunto ($\delta(x) \in \Omega$) (Priest, 1995/2002, p. 134).

La primera afirmación se denomina *Existencia*, las condiciones de la segunda afirmación se denominan *Transcendencia* y *Clausura* respectivamente. Estas son propiedades de cualquier conjunto bien fundado, y por lo tanto son una verdad lógica (Priest, 1995/2002, p. 130; Landini, 2009, pp. 106-107), pero detengámonos un momento en explicar las particularidades del esquema.[7]

Lo que propone el esquema es que el conjunto Ω de todos los objetos que cumplen la propiedad ϕ existe y tiene la propiedad ψ. Si se toma un subconjunto x de Ω, ϕ será una propiedad de todos los miembros de x, $\delta(x)$ tiene la propiedad ϕ y por lo tanto pertenece de Ω, pero no pertenece a x. Una instancia del esquema nos puede ayudar a clarificar esto.

Sea $\phi(y)$ la propiedad tal que 'y es par' y $\psi(\Omega)$ la propiedad 'Ω es igual a sí mismo'; Ω es el conjunto de números pares; x, en este caso, es el conjunto constituido por de los

---

B. Russell, quien pretendía demostrar que las paradojas de Russell-Zermelo, Cantor y Burali-Forti se reducían a una contradicción estrictamente lógica y que por lo tanto estas paradojas se resuelven cambiando algunos supuestos lógicos (Russell, 1906b, pp. 29-37) (previamente Poincaré ya había hecho comentarios similares al respecto (1906, p. 305)). Lamentablemente, Russell no fue capaz de explicar qué exactamente es lo que se debe cambiar, pues si bien presenta tres alternativas posibles, él mismo no se decide por ninguna al notar que las tres tienen amplias desventajas. Posteriormente, en *Principia Mathematica*, Whitehead & Russell propusieron algunas observaciones sobre cómo evitar los denominados *círculos viciosos*, pero en últimas estas han sido consideradas como poco satisfactorias. Ver (Gödel, 1944; Rouilhan, 1992; Priest, 1995/2002, cap. 9.6-9.8).

[7] Un conjunto está bien fundado cuando tiene un límite inferior, es decir, posee un primer elemento o un elemento mínimo.

pares mayores o iguales a 2 y menores o iguales a 6, por lo tanto $x = (2, 4, 6)$; y $\delta(x)$ es la función cúbica ($\delta(x) = x^3$). Aplicando $\delta(x)$ se obtiene el conjunto constituido por 8, 64 y 216, es claro que este conjunto no pertenece a x, pues aunque son pares, no son pares mayores o iguales a 2 y menores o iguales a 6, pero son pares al fin y al cabo, por lo que $\delta(x)$ sí pertenece a Ω.[8] Si bien esta instancia tiene que ver con números naturales, se pueden dar multitud de ejemplos que no necesariamente tengan que ver con estos, como se ve a continuación.

A pesar de lo claro que es el esquema, en este encajan una serie de paradojas. Es decir, el *Inclosure Schema* es una verdad lógica y permite contradicciones, ahí es donde yace lo paradójico. Uno de estos casos es la famosa paradoja de Russell-Zermelo, según la cual el conjunto de todos los conjuntos que no pertenecen a sí mismos, pertenece y no pertenece a sí mismo.

Esta paradoja encajaría en el esquema cuando se aplican los elementos del esquema de la siguiente manera: sea $\phi(y) = y \notin y$, $\psi(x) =$ 'x es igual a sí mismo', $\Omega = \{y : \phi(y)\}$, y $\delta(x) = x$; dado que $x \subseteq \Omega$, cualquier conjunto que pertenezca al conjunto de conjuntos que no pertenecen a sí mismos no pertenece a sí mismo (es decir, $\delta(x) \notin x$): cuando aplico $\delta(x)$, esta me arroja x de nuevo, y dado que x es un conjunto que no pertenece a sí mismo, es decir, $\phi(x)$ se cumple, no puede pertenecer a x. Aun así, cualquier conjunto que no pertenezca a sí mismo pertenece al conjunto de conjuntos que no pertenecen a sí mismos ($\delta(x) \in \Omega$), pues, de nuevo, $\delta(x)$ me arroja x, y ya se estableció que $\phi(x)$ se cumple, por lo tanto, x pertenece al conjunto de todos los conjuntos que no pertenecen a sí mismos. El problema está en que Ω es un conjunto que pertenece al conjunto de conjuntos que no pertenecen a sí mismos y no pertenece al conjunto de conjuntos que no pertenecen a sí mismos, pues dado $\Omega = \{y : \phi(y)\}$, para $\delta(\Omega)$ se tiene tanto que $\Omega \notin \Omega$ como que $\Omega \in \Omega$.

El mismo procedimiento puede aplicarse con una serie de paradojas identificadas por Priest en *Beyond the Limits of Thought* obteniendo el mismo resultado siempre: contradicciones que encajan en el *Inclosure Schema*, contradicciones a partir de verdades lógicas. Según Priest, las paradojas que encajan en el esquema son las de Russell-Zermelo, Burali-

[8] Puesto formalmente: Dado un conjunto Ω tal que $\Omega = \{y : \phi(y)\}$ donde $\phi(y) = y \in 2\mathbb{Z}$; dado que $x = \{z : \phi(z) \wedge 2 \leq z \leq 6\}$, y que $\delta(x) = x^3$; se obtiene $\delta(x) \notin x$ y $\delta(x) \in \Omega$.

Forti, Mirimanoff, la Quinta Antinomia de Kant,[9] König, Berry, Richard, la paradoja del mentiroso, la paradoja del conocedor (*knower paradox*) y la de Grelling-Nelson.[10] De esta manera, si bien todas estas paradojas encajan en el esquema, el que encajen no las trivializa, pues siguen siendo paradojas distintas dado que en todos los casos cambia Ω y cambia δ, como se ve a continuación.

Explicar cómo cada una de estas paradojas funciona y cómo encajan dentro del esquema no es objetivo principal de este trabajo. En este capítulo incluyo un cuadro en el cual expongo los valores que Ω, ϕ, ψ y δ deben tomar en cada caso (ver Cuadro 2.1). De igual manera, si el lector está particularmente interesado en alguna de las paradojas, lo remito al Apéndice A, donde se brinda una explicación detallada de cada paradoja.

Del Cuadro 2.1, sin embargo, hay un comentario importante que se debe hacer, que tiene relevancia para todo el desarrollo de este trabajo. Las paradojas referidas hasta el momento, a pesar de que tienen la misma estructura, no son exactamente la misma paradoja. Esto se debe que son sobre un conjunto de cosas distinto (Ω) y la función que permite autorreferencia es distinta (δ(x)).[11]

Por ejemplo, por más que en las paradojas de Burali-Forti y la de König los números ordinales sean relevantes, estas no son la misma paradoja, pues la primera es sobre el conjunto de números ordinales y la paradoja aparece al intentar encontrar el menor ordinal mayor que x, mientras que la segunda es sobre el conjunto de números ordinales *definibles*

---

[9] Este es el nombre que Priest utiliza para referirse a una paradoja que él mismo crea basado en las cuatro antinomias de Kant. Sin embargo, esta misma paradoja también había sido notada previamente por Merleau-Ponty, ver (1945/1994, pp. 13-14).

[10] Priest ha discutido si la paradoja de Curry también lo hace, su respuesta es que sí siempre y cuando se interprete el condicional como un condicional material. En años posteriores Priest ha demostrado cómo otras paradojas encajan dentro del *Inclosure Schema*, como la de Platón (ver los cadáveres era tan horrible que no podía quitarle los ojos de encima (República, 439a))(Priest, 2019) e incluso la paradoja de Sorites (Priest, 2010).

[11] El nombre que Priest le otorga a la función δ originalmente es diagonalizador. Para evitar confusiones con la diagonalización, que es un diagonalizador, prefiero referirme a los diagonalizadores como un tipo de generadores de autorreferencia o un tipo de mecanismo para lograr autorreferencia. Como se ve en el capítulo 4, existen otros mecanismos de autorreferencia en los lenguajes naturales; como se ve en el capítulo 5, el diagonalizador particular para la paradoja del mentiroso no es capaz de capturar las peculiaridades propias de algunos mecanismos de autorreferencia en lenguajes naturales.

Paradoja	$\phi(y)$ **Propiedad de los elementos del conjunto**	$\psi(x)$ **Propiedad del conjunto**	$\delta(x)$ **Función que permite autorreferencia**	Ω **Conjunto**
Russell-Zermelo	y no pertenece a sí mismo	x es igual a sí mismo	Función de identidad ($id(x)$)	R, el conjunto de conjuntos que no pertenecen a sí mismos
Burali-Forti	y es un ordinal	x es igual a sí mismo	Menor ordinal mayor que x ($log(x)$)	On, el conjunto de todos los ordinales
Mirimanoff	y es un conjunto bien fundado	x es igual a sí mismo	Conjunto potencia de x ($P(x)$)	V, el conjunto de conjuntos bien fundados
Quinta Antinomia	y pertenece al conjunto de pensamientos	x es igual a sí mismo	Pensar en x ($t(x)$)	T, el conjunto de todos los pensamientos
König	y es un ordinal	x es definible	Primer ordinal no definible ($\mu yy \notin x$)	DOn, el conjunto de ordinales definibles
Berry	y es un ordinal definible en menos de 19 palabras	x es definible en menos de 19 palabras	Primer ordinal no definible en menos de 19 palabras ($\mu yy \notin x$)	DN_{99}, conjunto de ordinales definibles en menos de 19 palabras
Richard	y es un número real entre 0 y 1 y es definible	x es definible	Primer número real definible de x ($diag(x)$)	DR, el conjunto de números reales definibles entre 0 y 1
Mentiroso	y es verdadera	x es definible	Una oración α tal que $\alpha = <\alpha \notin x>$	Tr, el conjunto de oraciones verdaderas
Conocedor (knower)	y es una oración conocida	x es definible	Una oración al tal que $\alpha = <\alpha \notin x>$	Kn, el conjunto de oraciones conocidas
Grelling-Nelson	y es un predicado que no aplica a sí mismo	x es igual a sí mismo	Propiedad de pertencer a x ($<v \in x>$)	Het, conjunto de predicados que no aplican a sí mismos

Cuadro 2.1: Paradojas autorreferenciales que cumplen con el *Inclosure Schema*. Fuente: (Priest, 1995/2002, pp. 131, 134, 146).

y la paradoja se genera cuando se intenta encontrar el primer ordinal no definible. De esta manera, dos paradojas tienen la misma estructura si cumplen con el esquema (en este caso, el *Inclosure Schema*), pero dos paradojas que cumplen el mismo esquema son distintas si son sobre un conjunto de elementos distinto y obtienen autorreferencia mediante funciones distintas.

De manera que por más que las paradojas de la autorreferencia sean similares en un nivel de abstracción y que haya un criterio para determinar si dos paradojas tienen la misma estructura, v.gr. el *Inclosure Schema*, también hay un criterio o principio de individuación entre las mismas: a pesar de que dos paradojas compartan una estructura particular, son diferentes paradojas si son respecto objetos distintos (Ω) y se generan de manera distintas (δ). Esto último significa que la función que se aplica en cada caso es distinta: en estas paradojas la fórmula

2.2 Sobre el Principio de Solución Uniforme

El análisis de las paradojas de la autorreferencia en términos del *Inclosure Schema* lleva a Priest a proponer lo siguiente: dado que todas estas paradojas tienen la misma estructura, deberían tener una solución uniforme. Es decir, al mismo tipo de paradoja, el mismo tipo de solución. Si dos paradojas fueran distintas, es de esperarse que estas tengan soluciones diferentes.

Este razonamiento se puede extender: si dos paradojas son de un tipo o clase distinta, entonces tienen soluciones de distinto tipo o clase; naturalmente, si dos paradojas son de la misma clase, se espera que tenga la misma clase de solución (Priest, 1995/2002, p. 183).[12] Esto constituye el Principio de Solución Uniforme (PUS). Este depende en gran medida de lo que se entienda por *tipo* o *clase* en ambos lados del condicional. Para que dos o más paradojas sean de la misma *clase* debe existir (a) una estructura que sea común a todas a todas las paradojas; y (b) esta estructura es la que produce la contradicción en todos los casos (Priest, 1995/2002, p. 183).

Así, si bien Priest fue el primero en analizar el rol que juega la autorreferencia tanto

[12] En el inglés original: "same kind of paradox, same kind of solution".

en la formulación de las paradojas como en la solución de las mismas: todas las paradojas de la autorreferencia comparten una misma estructura, es decir, hay un esquema según el cual, una vez reemplazados los valores relevantes, este es capaz de representar cualquier paradoja que contenga nociones autorreferenciales (Priest, 1994; Priest, 1995/2002, pp. 141-149; Priest, 2000, p. 123).

Volviendo al PUS, vale la pena detenerse en el sentido de *tipo* o *clase* en Priest cuando habla de "el mismo tipo/clase" de solución, es decir, de una solución uniforme. El punto sobre el que Priest quiere hacer énfasis es que debería existir un tipo o clase de solución que aplique a todas las paradojas. Las soluciones ortodoxas a las paradojas de Russell-Zermelo, la de Burali-Forti y la de Mirimanoff, i.e. construcciones de teorías axiomáticas de conjuntos, no son soluciones satisfactorias porque, entre otras cosas, no solucionan las paradojas de König, Richard, Berry, ni la del mentiroso, la del conocedor o la de Grelling-Nelson por el simple hecho que estas no son paradojas propias de la teoría de conjuntos (Priest, 1995/2002, p. 183). Se supone que, en un nivel de abstracción son la misma paradoja (cumplen con el *Inclosure Schema*), por lo tanto, deberían tener, en un nivel de abstracción, la misma solución (Priest, 2000, p. 125).

De modo similar, cuando se busca una solución uniforme a las paradojas del mentiroso debe tenerse en cuenta que dos soluciones incompatibles, por supuesto, no constituyen una solución uniforme, o del mismo tipo, pero que dos sean compatibles o coherentes entre sí no asegura una solución uniforme. Supongamos que tengo dos soluciones para distintas paradojas, por ejemplo, construyo una teoría axiomática de conjuntos para solucionar varias de las paradojas citadas, construyo una jerarquía de lenguajes que me permita solucionar la paradoja del mentiroso. Estas soluciones son coherentes, pues no se contradicen entre sí. Aún así, dice Priest, esto no sería cumplir con el PUS, pues a pesar de que las soluciones funcionen individualmente, estas soluciones no comparten nada entre sí, por lo que no son del mismo tipo (Priest, 2000, p. 124). El hecho que dos soluciones a paradojas sean coherentes, no implica que de manera colectiva constituyan la misma clase de solución.

Así, una solución uniforme, según Priest, es una que resuelva las contradicciones subyacentes a todas las paradojas de la autorreferencia. La aproximación que logra esta uniformidad de manera satisfactoria, según él, es el dialeteísmo, pues el dialeteísmo afirma que la solución consiste en aceptar que no hay un problema con que las paradojas sean afirmaciones contradictorias, ya que son dialetheiai, es decir, son verdaderas y falsas al mismo tiempo y en el mismo sentido (Priest, 1995/2002, p. 169).

Las implicaciones de esto son diversas y se ha escrito mucho en pro y en contra del dialeteísmo. Es importante resaltar que en tanto esta ha sido una explicación de la motivación del dialeteísmo de Priest, no se pretende ni defender ni atacar esta posición. Para una defensa del dialeteísmo de Priest, sugiero Priest (1987/2006; 2006); para revisar la que tal vez es una de las críticas más contundentes al dialeteísmo (v.gr. cómo la paracompletitud es una solución uniforme mejor que el dialeteísmo), sugiero Field (2008, ver especialmente sec. 25).

2.3 El lugar de la paradoja del mentiroso en el *Inclosure Schema*

La paradoja del mentiroso ocupa un lugar especial en la historia de la filosofía occidental, no solo porque es de las más antiguas, sino porque es una de las más estudiadas. Sin embargo, cuando se habla de 'paradoja del mentiroso' no siempre es claro a qué paradoja nos estamos refiriendo. Esto se debe a que a lo largo de la historia se le ha llamado paradoja del mentiroso a una serie de contradicciones que surgen cuando se pretende atribuir falsedad de manera autorreferencial.

En *Beyond the Limits of Thought*, Priest presenta de alguna u otra manera hasta cinco formulaciones de la paradoja del mentiroso. Dentro de estas, a las que más les dedica atención es a las siguientes tres:

- Estoy mintiendo (*I am lying* (1995/2002, pp. 142-143)), y
- Esta oración es falsa (*This sentence is false* (1995/2002, pp. 144, 290)),
- $\sigma = \langle \sigma \notin Tr \rangle$ (1995/2002, pp. 144, 290).[13]

[13] Una de las labores más arduas de este trabajo ha sido uniformizar la notación. De antemano me disculpo si la notación que aparece en alguno de los pasajes no concuerda con la de los textos originales, pero esto ha sido con el afán de presentar un trabajo con una notación consistente.

A pesar de que estas tres formulaciones aparecen a lo largo del libro, Priest solo demuestra cómo la tercera de estas ($\sigma = \langle \sigma \notin Tr \rangle$) encaja dentro del *Inclosure Schema*. Esto se debe, en gran parte y como se explica un poco más adelante, a que Priest piensa que estas tres son exactamente la misma paradoja, por el momento, veamos cómo Priest encaja la versión de la paradoja en lenguaje formal en el *Inclosure Schema*.

Asumiendo que las oraciones son las portadoras de la verdad, y siguiendo los valores presentados en el cuadro 2.1, podemos construir la paradoja al estilo de Priest: $\phi(y)$ es tal que 'y es verdadera', de modo que Ω es el conjunto de oraciones verdaderas, Tr; $\psi(x)$, por su lado, es tal que 'x es definible', $\delta(x)$ es una función $\sigma(x)$ definida por un método de diagonalización[14] tal que si a es un conjunto y satisface la propiedad ψ tal que $\psi(a)$, entonces $\sigma(a) = \alpha$, donde $\alpha =< \alpha \notin a >$, de tal manera que que la expresión en corchetes es una oración que expresa el hecho que α no es un elemento de a. Nótese que en este caso α es decir el nombre de a. Este fragmento se trata con más detalle en el capítulo 3.

Recuérdese que el *Inclosure Schema* está formulado en los siguientes términos:

1. Dado un conjunto Ω, tal que $\Omega = \{y : \phi(y)\}$ y $\psi(\Omega)$ (Existencia)
2. Si $x \subseteq \Omega$ y $\psi(x)$:
   (a) $\delta(x) \notin x$ (Transcendencia)
   (b) $\delta(x) \in \Omega$ (Clausura)

Ahora, suponiendo que a es definible y $a \subseteq Tr$, demostrar Trascendencia y Clausura es el siguiente paso. Para simplificar las cosas, (1) $\delta(x)$ será simplemente σ y (2) $\sigma = \langle \sigma \notin a \rangle$, donde σ sigue siendo el nombre a.[15]

[14] Una explicación detenida y en detalle de la función de diagonalización se da más adelante, sin embargo, por ahora basta con entender a grandes rasgos en qué consiste. En pocas palabras, diagonalizar es tomar una variable libre de una fórmula y reemplazarla por el nombre de esa fórmula, de manera que la diagonalización de $B(x)$ es $B(\ulcorner B(x) \urcorner)$ ($\ulcorner n \urcorner$ representa el nombre de n), de forma que la diagonalización de la fórmula lo que dice que la fórmula misma (el nombre) satisface la propiedad que la fórmula designa.

[15] Si bien esta no es la formulación de Priest en (1995/2002), es un simplificación que presentó posteriormente (2012, p. 368).

σ precisamente me arroja α (el nombre de a, también representado mediante $\ulcorner a \urcorner$), una oración que dice que α no pertenece a al conjunto a, por lo tanto $\sigma(a) \notin a$, se cumple Trascendencia. Al mismo, tiempo, a partir de σ, obtengo $\alpha \notin a$, y, por esquema-T, $< \alpha \notin a > \in$ Tr. Dado que $\sigma(a) = \alpha \notin a$, entonces $\sigma(a) \in$ Tr y se cumple Clausura. La paradoja del mentiroso es la oración $\sigma(Tr)$: dado que $\sigma(a) =< \alpha \notin a >$, entonces $\sigma(Tr) =< \ulcorner Tr \urcorner \notin Tr >$. Dado que Trascendencia y Clausura, $\sigma(Tr) \notin Tr$ y $\sigma(Tr) \in Tr$.

Pero aquí es cuando Priest comienza a saltar entre paradojas como si fueran equivalentes. Previamente, discutiendo un fragmento de Ramsey (1925, p. 20), Priest se refiere a la oración 'estoy mintiendo' (*I am lying*) como la paradoja del mentiroso (Priest, 1995/2002, p. 143). Pero una página después, cuando explica cómo afirma que la paradoja del mentiroso encaja en el *Inclosure Schema*, Priest construye una versión de la paradoja en términos formales que es la que se acaba de detallar, pero a su vez afirma que esta formulación de la paradoja es una construcción que dice 'esta oración es falsa':

> Let ϕ(y) be 'y is true', so that Ω is the set of true sentences, Tr; let ψ(x) be 'x is definable'. δ is a function, σ, defined by some suitable technique of diagonalisation so that if a is any definable set σ(a) = α where, $\alpha = \langle \alpha \notin$ a $\rangle$. The angle-bracket expression is a sentence expressing the fact that α is not in the set a. (Hence, α says '**This sentence is not in** *a*'.) (Priest, 1995/2002, p. 144, énfasis mío)

Lo que Priest está diciendo en este fragmento es que la paradoja del mentiroso que encaja en el *Inclosure Schema* se genera gracias al uso de la diagonalización sobre el predicado de no pertenencia. Cuando se construye una oración mediante diagonalización tal como se indica, esta oración lo que dice es '*esta* oración no está en *a*'.

En este sentido, el primer salto que da Priest es del lenguaje natural al lenguaje formal. Dice que la paradoja del mentiroso es 'estoy mintiendo', pero pasa a construir la paradoja en términos de $\sigma = \langle \sigma \notin Tr \rangle$. Pero eso no es todo, pues, como acabamos de ver, Priest toma esta formulación en lenguaje formal y vuelve al lenguaje natural: según él, debemos interpretar $\sigma = \langle \sigma \notin Tr \rangle$ como 'esta oración no pertenece al conjunto de oraciones verdaderas' v.gr. 'esta oración es falsa'.

Estos saltos no son un mero desliz en la argumentación de Priest pues él no nos dice nada acerca de cómo las otras paradojas encajan en el *Inclosure Schema* y sin embargo dice: "All the contradictions with which the book concerns itself are inclosure contradictions, that is, contradictions fitting the Inclosure Schema. In particular, it is shown that all the standard paradoxes of self-reference fit the Schema" (Priest, 1995/2002, p. 277).

Los problemas se multiplican cuando, más adelante, comentando sobre las formulaciones de la paradoja del mentiroso, Priest dice:

> The Liar paradox comes in many varieties. In some of them, self-reference is achieved by means of a demonstrative ('this sentence', etc.); in others is obtained by a description ('the first sentence on such and such page', etc.); yet in others is obtained by some diagonal argument using Gödel-numbers as names. Now, at some level of abstraction, these paradoxes are, presumably, of different kinds. The first, after all, depends on context-dependent features of utterance; the second presupposes some account of the way that descriptions work; the third employs numbers. (...) [But] **The exact *mechanisms* of reference are, in a clear sense, accidental to what is going on**. (Priest, 1995/2002, 290, la cursiva es del original, la negrilla es mía)

De estas dos afirmaciones debería seguirse que todas las paradojas del mentiroso encajan en el *Inclosure Schema*. Pero Priest solo explica cómo la versión de la paradoja en lenguaje formal encaja en el esquema. Considerando esto a la luz de las dos afirmaciones de Priest que se acaban de exponer es claro que para él no hay una diferencia entre la formulación de la paradoja en lenguaje formal y las formulaciones en lenguaje natural entre las que va saltando ('estoy mintiendo' y 'esta oración es falsa'): en tanto ya demostró que la paradoja en lenguaje formal encaja en el esquema, no es necesario probar que las otras lo hacen, pues los mecanismos con los que se logra autorreferencia son accidentales, por lo tanto, todas las formulaciones de la paradoja del mentiroso encajan en el esquema.

Sin embargo, no creo que baste mostrar que la formulación en lenguaje formal cumple con el esquema para demostrar que todas las formulaciones de la paradoja cumplen con el esquema, pues esto sería aceptar que las formulaciones en lenguaje natural se reducen a la formulación en lenguaje formal. Es decir, esto es a aceptar que las formulaciones en lenguaje natural cumplen con el esquema en tanto la formulación en lenguaje lo hace: en tanto $\sigma = \langle \sigma \notin Tr \rangle$ cumple con el esquema, 'estoy mintiendo' lo hace; en tanto $\sigma = \langle \sigma \notin Tr \rangle$ cumple con el esquema, 'esta oración es falsa' lo hace.

Mi escepticismo frente a esto tiene que ver con lo referido anterioremente sobre cómo, a pesar de que varias paradojas tengan la misma estructura, siguen siendo paradojas distintas. Al final de sección 2.1 se planteó la siguiente condición que vincula el *Inclosure Schema* con el hecho que este agrupa paradojas que al final de cuentas son distintas: dos paradojas tienen la misma estructura si cumplen con el esquema (en este caso, el *Inclosure Schema*), pero dos paradojas que cumplen el mismo esquema son distintas si son sobre un conjunto de elementos distinto y obtienen la autorreferencia de manera distinta.

Así, cuando se analizan las paradojas, es evidente que hay algunas diferencias en cuanto a los elementos a los que refieren y en cuanto a la forma en la que obtienen autorreferencia. A saber, la versión de la paradoja en lenguaje formal habla sobre fórmulas de un lenguaje formal y logra autorreferencia mediante la diagonalización; mientras que en las formulaciones en lenguaje natural sucede lo siguiente: por un lado, 'estoy mintiendo' refiere a oraciones dichas en un contexto por un agente particular y logra autorreferencia mediante la palabra 'yo'; por el otro, 'esta oración es falsa' refiere a oraciones y logra autorreferencia mediante la palabra 'esta'.

Los capítulos restantes de este trabajo exploran si estas diferencias bastan para decir que en estos casos no estamos solo frente a diferentes formulaciones de la paradoja del mentiroso, sino ante genuinas paradojas del mentiroso. El capítulo 3 explora el concepto de diagonalización y el concepto de referencia en lenguajes formales; el capítulo 4 se encarga de examinar la referencia mediante 'yo' y 'esta'; por último, el capítulo 5 recoge lo dicho en los anteriores dos capítulos y da buenas razones para pensar que estas tres formulaciones son en realidad tres paradojas distintas.

Capítulo 3

Las paradojas del mentiroso en lenguajes formales: herramientas para su análisis

"*In arithmetic the differences don't matter, because in arithmetic nothing is contingent or a posteriori. Whatever satisfies a definite description in arithmetic does so by necessity and a priori. So all designators are rigid.*"
Fred, en el diálogo *The Root of Evil*
(Halbach, 2016, p. 159)

En el capítulo anterior se presentó la motivación del dialeteísmo de Priest y el lugar de la paradoja del mentiroso en esta. A su vez, se presentaron las tres versiones de la paradoja en las que Priest se detiene y se explicó cómo, según él, todas estas encajan en el *Inclosure Schema* en tanto la versión en lenguaje formal lo hace. El presente capítulo tiene como propósito presentar el aparato teórico que permite analizar la paradoja del mentiroso en lenguaje formal para determinar a qué hace referencia y cómo lo logra. Una vez más, la versión formal de la paradoja del mentiroso según Priest es esta (1995/2002, p. 144):

- $\sigma = \langle \sigma \notin \ulcorner Tr \urcorner \rangle$

Recuérdese que en la sección 2.3 se explicó cómo a partir de esta fórmula Priest deriva la contradicción propia de la paradoja del mentiroso.[16] En síntesis, $\sigma = \langle \sigma \notin \ulcorner Tr \urcorner \rangle$ representa una oración que dice *de sí misma* que es falsa (Priest, 1995/2002, p. 144). El objetivo de este capítulo es determinar qué significa que $\sigma = \langle \sigma \notin \ulcorner Tr \urcorner \rangle$ diga algo *de sí misma*.

Con tal fin, primero se explica cómo es posible obtener una oración como $\sigma = \langle \sigma \notin \ulcorner Tr \urcorner \rangle$ a través de una función representable en algunos sistemas formales llamada diagonalización. A la luz de esto, se explica en términos generales qué significa atribuirse a sí misma una propiedad a través de la diagonalización mediante un cuantificador existencial, que es la forma predilecta de Priest para obtener fórmulas autorreferentes.

3.1 Autorreferencia en lenguajes formales

Desde la publicación del revolucionario trabajo de Gödel (1931/1986) sobre la incompletitud de la aritmética se ha reconocido la posibilidad de que las funciones y los sistemas formales hablen de sí mismos. En su *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Gödel demostró (1) que en cualquier sistema formal que sea consistente, axiomatizable y que parta de la aritmética de Peano, existen proposiciones formalmente indecidibles (es decir, proposiciones que son verdad pero no hay un procedimiento finito para probar estas ni su negación) y (2) que la consistencia de un sistema formal con las características mencionadas no puede ser probada en el sistema mismo.

Así, a partir de Gödel es común hablar de autorreferencia en lenguajes formales. Pero mucho menos común es una explicación detallada de lo que significa que una oración de un lenguaje formal sea autorreferencial.[17] Esto es, no se explica, por ejemplo, cómo se logra

[16] Heck (2007, p. 5) obtiene la paradoja de manera similar a la de Priest. Sin embargo, es importante resaltar que esta no es la única formulación de la paradoja en lenguaje formal. Kripke (1975, pp. 690-1), por ejemplo, propone que para cualquier predicado 'Qx' tal que 'x es falso' puede producirse un predicado 'Px' tal que '$\forall x(Px \rightarrow Qx)$' es una formulación de la paradoja cuando'Px' es el predicado 'la cuarta oración entre comillas de la nota al pie de página 16 de la tesis de pregrado de Pablo Rivas-Robledo'. Por supuesto, Kripke usa un ejemplo distinto.

[17] Leitgeb (2002) es uno de los primeros en denunciar este hecho, recalcando la imposibilidad de dar cuenta de este fenómeno en términos extensionales. Halbach y Visser (2014a; 2014b) mantienen que

autorreferencia: si se logra mediante el uso de indéxicos, demostrativos, descripciones, etc. Respecto a esto, Smullyan ha sugerido que si bien es posible formalizar el uso de demostrativos, estos no ocurren en los sistemas estudiados por Gödel, por lo tanto no es posible afirmar que Gödel lograra sus autorreferencia por medio de demostrativos (1994, p. 2).[18]

Lo cierto es que Priest utiliza la misma herramienta que Gödel para lograr autorreferencia, la diagonalización (también llamada estrategia de puntos fijos) y por lo tanto el trabajo de Priest tiene las mismas limitaciones que el trabajo de Gödel.[19] Sin embargo, hay poco consenso respecto a cómo debe entenderse que una oración autorrefiera mediante diagonalización: algunos sugieren que se trata de una oración con indéxicos (Kleene, 1986, p. 134; Gaifman, 2006, p. 713; Raatikainen, 2018, sec. 4.3; Smullyan, 1992, p. 11) y esta parece ser un opinión generalizada (Milne, 2007, Appendix); otros afirman que se trata de una oración con demostrativos (Priest, 1995/2002); mientras que un último grupo tiene buenas razones para pensar que se trata de una descripción (Heck, 2007, p. 5; Halbach, 2016, p. 162).

Como se mencionó, Priest acepta que se trata de una oración con demostrativos:

> Let ϕ(y) be 'y is true', so that Ω is the set of true sentences, Tr; let ψ(x) be 'x is definable'. δ is a function, σ, defined by **some suitable technique of**

---

una explicación adecuada debe darse tomando en cuenta varios aspectos intensionales. Uno de los esfuerzos recientes más significativos por dar cuenta de este fenómeno para la aritmética de primer orden (y la aritmética de primer orden extendida con un predicado de verdad) es el de Picollo (2018; 2020a; 2020b).

18 Para una explicación de cómo y bajo qué condiciones es posible formalizar demostrativos para lograr autorreferencia, ver Smullyan (1984).

19 Debe advertirse que Gödel nunca utilizó los términos *diagonalización* o *puntos fijos*, estos fueron acuñados de manera posterior para referirse al método que utiliza Gödel en la introducción y en el teorema VI de (1931/1986) por la semejanza que tiene con todo de las diagonales de Cantor, con el cual probó que el conjunto de los números reales es de mayor cardinalidad que el de los naturales. Lo que hace Gödel en (1931/1986) es construir una oración similar a la pardoja del mentiroso, la cual expresa que no se puede demostrar en el sistema. Como posteriormente fue comprobado por Carnap, este procedimiento se puede aplicar para cualquier propiedad expresable en el sistema (1934/2001, §35), y por eso es que este resultado ha tomado el nombre de teorema de puntos fijos de Gödel-Carnap (Gaifman, 2006, p. 710).

diagonalisation so that if a is any definable set $\sigma(\mathrm{a}) = \alpha$ where, $\alpha = \langle \alpha \notin \mathrm{a} \rangle$. The angle-bracket expression is a sentence expressing the fact that α is not in the set a. (Hence, α says '**This sentence is not in a**'.) (Priest, 1995/2002, p. 144, énfasis mío)

Recuérdese además que, como se mencionó en el capítulo anterior, esta cita aparece después de que Priest haya identificado la paradoja del mentiroso como la afirmación 'estoy mintiendo'. De nuevo, este tipo de equivalencias es donde parece que yace el problema con el lugar que Priest le adjudica a la paradoja del mentiroso.

Sin embargo, creo que afirmar que se logra autorreferencia con demostrativos o con indéxicos a través de la diagonalización no solo es una afirmación injustificada en el texto de Priest, sino que es un error. No solo por el comentario de Smullyan mencionado más arriba, sino porque cuando se detalla qué significa obtener una oración mediante el método de diagonalización es posible darnos cuenta que es una fórmula cuyo cuantificador principal es un existencial afirmativo.[20] En lo que queda del capítulo justifico esta afirmación.

3.2 Referencia mediante cuantificadores

Como más adelante se hace explícito, la versión en lenguaje formal de la paradoja del mentiroso formulada por Priest es un enunciado de aritmética de primer orden de la forma $\exists x(\phi \wedge \psi)$ obtenido mediante diagonalización. Es decir, un enunciado cuyo operador principal es un cuantificador existencial positivo.

La naturaleza precisa de la autorreferencia mediante diagonalización y su relación con otros mecanismos de autorreferencia en lenguajes naturales ha sido poco explorada. Heck (2007, p. 5) asegura que un punto fijo logra autorreferencia mediante descripción (*by description*), pero no da explicaciones al respecto. Picollo (2018; 2020a) en cambio la llama referencia por cuantificación o referencia-q (*q-reference*) y explica las condiciones bajo las cuales hay referencia-q.

[20] En lo que sigue tomo la denominada construcción existencial del lema de diagonalización o diagonalización$^{\exists}$, consciente de que existe otro modo de lograr el mismo resultado mediante una construcción universal del lema de diagonalización, ver Milne (2007, sec. 3).

La tesis de Heck parece ser convincente: un cuantificador existencial como $\exists x P(x)$ normalmente se lee como 'existe un x tal que x tiene la propiedad P', la cual comúnmente se acepta que es la forma lógica que tienen las descripciones, que fue formulada por primera vez por Whitehead y Russell en *Principia Mathematica* (Whitehead & Russell, 1910/1927, $*14$). Por lo cual resulta atractivo pensar que los puntos fijos obtenidos mediante el Lema de Diagonalización son descripciones. Y si bien el descriptivismo ha sido fuertemente criticado como una teoría del significado en la segunda mitad del siglo XX, el que las descripciones se puedan formalizar como enunciados particulares afirmativos es un punto que rara vez se pone en duda cuando se discute el descriptivismo.[21]

Pero una cosa es que una descripción se pueda formalizar como un existencial afirmativo y otra cosa es que un existencial afirmativo tenga la forma lógica de una descripción. Heck parece afirmar lo segundo, pero no parece ser siempre el caso: $\exists x(x = a \wedge P(x))$ afirma que a es tiene la propiedad P, pero no puede ser una descripción porque precisamente usa constantes para referir, las cuales no sirven para representar descripciones definidas (Bergmann y col., 1980/2014, sec. 7.5).

Teniendo esto en cuenta, si bien la paradoja del mentiroso en lenguaje formal es un existencial afirmativo, debe tomarse con cautela la afirmación de que los puntos fijos refieran por medio de descripciones.[22] De este modo, prefiero acercarme a la referencia en lenguajes formales mediante la propuesta de Picollo.[23]

21 Como es de esperarse, hay voces que difieren: Fara (2001; 2016) sostiene, por ejemplo, que las descripciones son predicados.

22 En la filosofía de las matemáticas el naturalismo platonizado de Linsky y Zalta se ha consolidado como otra postura según la cual la referencia a objetos matemáticos se da por medio de descripciones. Sin embargo, aceptar este aspecto de Linksy y Zalta conlleva aceptar una ontología de corte meinongniano muy particular con la que no me quiero comprometer. Para el naturalismo platonizado ver Linsky & Zalta (1995), para su teoría sobre la referencia a objetos matemáticos ver (1995, pp. 532-538; 2019). La ontología de corte meinongniano mencionada se encuentra principalmente en Zata (1983).

23 En concreto, Picollo ha desarrollado dos aparatos para explicar la referencia en lenguajes formales, uno para la aritmética de primer orden ($\mathcal{L}$) y otro para la aritmética de primer orden extendida con el predicado de verdad ($\mathcal{L}_{\mathcal{T}}$). En este escrito trabajo sobre el primero de estos dado que para obtener la paradoja al estilo de Priest no se necesita extender el lenguaje con un predicado de verdad, dado que según Priest el predicado de verdad se puede definir recursivamente en el sistema

Entender adecuadamente el concepto de autorreferencia depende de entender adecuadamente el concepto de referencia. Una vez se tiene el segundo, basta con limitar el domino de objetos a los que una oración se refiere para lograr el concepto de autorreferencia. Sin embargo, las discusiones en torno a la naturaleza de la autorreferencia en lenguajes formales en parte responde a un entendimiento defectuoso de la noción de autorreferencia en lenguajes formales, según la cual una oración refiere a sí misma si es lógicamente equivalente a una oración que le atribuye una propiedad ($\phi \leftrightarrow B(\psi)$) que de hecho ha plagado la discusión en torno a si la paradoja de Visser-Yablo es autorreferente (Leitgeb, 2002, sec.s 2-3; Picollo, 2020a, sec. 2.3).

Si esto fuese así, cualquier instancia del esquema T de Tarski de la forma $Tr(\ulcorner\psi\urcorner) \leftrightarrow \psi$ sería una oración autorreferencial, lo cual no es cierto (Halbach & Visser, 2014b, secs. 7.2-7.3; Picollo, 2018, p. 579). Es así como Picollo introduce criterios que debe cumplir cualquier noción precisa y efectiva de referencia e introduce dos maneras en las que puede haber referencia en la aritmética de primer orden que cumplen con estos desiderata, referencia-m y referencia-q (Picollo, 2018, § 4). Esto obedece a que en la aritmética de primer orden las fórmulas bien formadas o son fórmulas atómicas, combinaciones de estas mediante conectivos lógicos o son fórmulas que contienen cuantificadores sobre fórmulas atómicas o combinaciones de estas mediante conectivos lógicos.

La referencia-m o referencia por mención, se da entre dos fórmulas ϕ y ψ, sii (si y solo si) ϕ contiene un término t tal que en la interpretación establecida $t = \ulcorner\psi\urcorner$, donde $\ulcorner\psi\urcorner$ representa el nombre de la fórmula ψ. Es importante notar que en este caso t no solo es lógicamente equivalente a $\ulcorner\psi\urcorner$ ($t \leftrightarrow \ulcorner\psi\urcorner$), sino equivalente por igualdad a $\ulcorner\psi\urcorner$ ($t = \ulcorner\psi\urcorner$). Esta es una condición más estricta pensada para superar el problema que suponía interpretar las instancias del esquema T como autorreferenciales. Así, por ejemplo, bajo una interpretación establecida, la expresión $B(a)$, donde $B(x)$ denota una propiedad y a es el nombre de una fórmula $0 = 0$, la expresión $B(a)$ refiere a $0 = 0$.

Dar cuenta de la referencia-q es un tanto más complicado, pero el trabajo de Picollo captura y extiende la idea según la cual expresiones de la forma $\forall x(\phi(x) \rightarrow \psi(x))$ y $\exists x(\phi(x) \wedge \psi(x))$ refieren a todo(algún) x que satisfaga ϕ y también refieren a todo(algún)

Q (que presento en la siguiente sección) a partir del predicado de comprobabilidad, dado que todo lo que es comprobable en el sistema es verdad en este (Priest, 2007, p. 186).

x que satisfaga ψ siempre y cuando este satisfaga ϕ (Picollo, 2018, p. 584). Por ejemplo, si $\phi(x)$ expresa que x es hombre y que $\psi(x)$ es mortal, $\forall x(\phi(x) \to \psi(x))$ se refiere a todos los x que satisfacen tanto ϕ (son hombres) como ψ (son mortales).

Sin embargo, no todas las expresiones son tan simples, pues pueden tener cuantificadores aglutinados o predicados no-monádicos. Para poder aplicar estos criterios a oraciones más complejas, el primer paso es obtener su forma normal postnexa disyuntiva. Esto es, para una fórmula ϕ, encontrar una forma de expresar la misma fórmula (1) sin hacer uso de los conectores $\to$, $\leftrightarrow$, el cuantificador $\exists$; (2) eliminando los cuantificadores superfluos; (3) transformando todas las subfórmulas de ϕ tal que sean disyunciones de conjunciones de fórmulas atómicas, negaciones de fórmulas atómicas, fórmulas con cuantificador universal o negaciones de estas; (4) introduciendo los cuantificadores lo más que se pueda dentro de la fórmula y (5) que dentro de cada subfórmula de la forma $\forall v_1 ... \forall v_n(\phi_1 \vee ... \vee \phi_m)$, v_n ocurre libremente en ϕ_j.

Por ejemplo, la fórmula normal postnexa disyuntiva de la fórmula $\forall x(\forall y \exists z \neg((x = 0 \vee z = 0) \wedge x = y))$ es la fórmula $\forall x(\neg \forall z \neg(\neg(x = 0) \wedge \neg(z = 0)) \vee \neg \forall y(x = y))$. De esta manera, la nueva fórmula se obtiene de ϕ aplicando las transformaciones referidas anteriormente y la nueva fórmula será lógicamente equivalente a cualquier transformaciones que impliquen la sustitución de variables ligadas; interdefiniciones lógicas; las Leyes de De Morgan; contraposición y distribución de la conjunción sobre una disyunción (y viceversa); la conmutatividad de la conjunción, disyunción, identidad y de los cuantificadores; la asociatividad de conjunciones y disyunciones; la introducción de cuantificadores superfluos; y la introducción y eliminación de negaciones dobles.

De esta manera, referencia-q se define de la siguiente manera: dada una fórmula en forma normal postnexa disyuntiva y una interpretación particular se cumple lo siguiente (Picollo, 2020a, Def. 3; Picollo, 2018, pp. 594-595):

1. $\forall v \phi$ q-refiere a todo objeto que satisfaga ϕ dada la interpretación establecida, siempre y cuando ϕ sea una expresión atómica, una negación de una expresión atómica o una conjunción.

2. $\forall v(\phi \vee \psi)$ [i.e. $\forall v(\neg \phi \to \psi)$] q-refiere a todo lo que satisfaga $\neg \phi$ dada la interpreta-

ción establecida y a lo que ψ q-refiera o m-refiera.[24]

Así, por ejemplo, la fórmula $\forall x(x = 1 \rightarrow \forall y(y = x + 2 \rightarrow Impar(y)))$ q-refiere a (todo objeto que sea igual a) 1 y a lo que la subfórmula $\forall y(y = x + 2 \rightarrow Impar(y))$ q-refiere, a (todo objeto que sea igual a) 3.

Con esto en mente, una fórmula ϕ refiere directamente a una fórmula ψ sii o m-refiere a ψ o q-refiere a ψ. Una fórmula también puede referir mediante una cadena de referencia si es parte de una secuencia de fórmulas tal que todas las fórmulas refieren directamente a la fórmula siguiente. Así, la referencia para la aritmética de primer orden se define de la siguiente manera: una fórmula ϕ refiere a una fórmula ψ sii hay una cadena de referencia que comienza en ϕ y termina en ψ. De esta manera, la autorreferencia es el fenómeno que se presenta cuando una fórmula ϕ se refiere a sí misma, es decir, si hay una cadena de referencia que comienza en ϕ y termina en ϕ (Picollo, 2020b, Defs. 4-7).

3.3 Obteniendo $\sigma = \langle \sigma \notin Tr \rangle$

Contrario a lo que sucede con los lenguajes naturales, obtener la paradoja del mentiroso es significativamente más complejo en lenguajes formales. Esto se debe a que el lenguaje que utilizan tanto Priest como Gödel para lograr autorreferencia está diseñado para expresar la aritmética. Sin embargo, si tomamos todas las expresiones del sistema y las ponemos en una manera completamente sintáctica, de modo que un número natural sea el nombre sintáctico de una expresión y repetimos este proceso para todas las expresiones del sistema, se podría lograr autorreferencia en el sistema en la medida que las expresiones del sistema expresen propiedades acerca del número que es su nombre sintáctico.

El método que se utilizará (la diagonalización) toma las expresiones del sistema y realiza lo siguiente (ver Cuadro 3.1):

Ahora bien, hay una variedad de formas de obtener el lema de diagonalización, pues el método de Gödel se ha simplificado y mejorado varias veces a lo largo de los años. Aquí,

[24] Esta manera particular de definir la referencia-q fueron tomadas de un documento fechado del 31 de enero de 2020 elaborados por Lavinia Picollo y que me hizo llegar por medio electrónico.

Expresiones de un sistema que expresa la aritmética	→	Asignarle a cada expresión un nombre, que será un número natural
↑		↓
Y dado que el sistema puede representar las nociones del sistema, puede representar	←	Todas las nociones del sistema, como prueba, axioma, etc., estarán asociadas a nociones aritméticas del sistema

Cuadro 3.1: Síntesis del método para lograr autorreferencia mediante diagonalización. Inspirado en (Berto, 2009, p. 85).

sin embargo, procuro construir el lema tal como Priest lo logra en *In Contradiction*, donde la sección del tercer capítulo incluye una prueba del primer teorema de la incompletitud de Gödel, del cual hace parte una construcción del lema de diagonalización (1987/2006, cap. 3.5).

3.3.1 Diagonalización: una primera aproximación

El lema de diagonalización o teorema de puntos fijos es un resultado comprobabale en algunos sistemas formales que en parte depende de que todas las funciones del sistema sean *nombrables* es decir, que se les pueda asignar un nombre de manera efectiva (esto es, que dado el nombre de la función, hay una forma de determinar qué expresión nombra y viceversa (Heck, 2007, p. 22; Halbach & Visser, 2014a, p. 674)). Esto es posible mediante una serie de métodos pero la más famosa es la numeración de Gödel, la cual también fue desarrollada en (1931/1986).

La numeración gödeliana busca asignarle un nombre único a todas las expresiones de un sistema formal asignándole un número a cada símbolo de cada expresión. Por ejemplo, el nombre de la expresión $(0 = 0)$ sería la serie de números que se forman concatenando el número para '(', el número para '0', el número para '=', el número para '0'y el número para ')'. Supongamos que utilizamos el siguiente esquema de numeración (ver Cuadro 3.2):

Según este esquema, el nombre de la expresión $(0 = 0)$ sería el numeral **174719**, esto

Símbolo	(	)	=	0
Código	1	19	4	7

Cuadro 3.2: Un ejemplo de numeración gödeliana.

es el número **0** seguido de 174719 signos sucesor ‘’’.[25] El numeral de una expresión ϕ del sistema se puede representar como $\ulcorner\phi\urcorner$. Así, $\ulcorner(0 = 0)\urcorner = \mathbf{174719}$. Por supuesto, el esquema debe ser ampliado de modo que a cada símbolo del sistema le corresponda un código.[26]

Aunque más adelante se profundiza en esto, vale la pena anotar que diagonalizar consiste en tomar una fórmula con una variable libre y aplicarle la fórmula al nombre de la fórmula (Gaifman, 2006, p. 710). Esto es, si se tiene una fórmula $B(x)$ donde x es una variable libre, la diagonalización de $B(x)$ sería $\exists x(x = \ulcorner B(x)\urcorner \wedge B(x))$, que es equivalente a $B(\ulcorner B(x)\urcorner)$ (Boolos y col., 1974/2007, p. 220). Esto último significa que la diagonalización es una función del sistema que se puede utilizar una vez se tenga el mecanismo adecuado para nombrar todas las fórmulas del sistema. Es momento entonces de especificar qué sistemas lo permiten.

3.3.2 Obteniendo el lema de diagonalización: los requisitos del sistema

En esta sección utilizaremos el sistema **Q** o aritmética de Robinson, que contiene los axiomas mínimos necesarios para obtener el lema de diagonalización (Mostowski y col., 1953).[27] En este se pueden definir las operaciones de suma y multiplicación y la función

[25] En el sistema que utilizaremos se parte del número 0 y de ahí cualquier número natural se puede obtenerse mediante la operación de sucesión, como se detalla más adelante. De modo que, por ejmplo, **2** es el *numeral* del número 0”. Se utiliza la negrilla para distinguir cuando un número natural que nombra una expresión de los números naturales que simplemente se obtienen a partir de la función de sucesor.

[26] Vale la pena que el esquema de numeración presentado aquí no está completo. Para revisar el esquema completo ver (Boolos y col., 1974/2007, p. 188).

[27] Si bien es cierto que hay sistemas más débiles en los cuales se pueden lograr resultados muy similares usando la función de normalización, aquí me enfoco en la diagonalización por ser el método de preferencia de Priest. Para consultar más sobre otros métodos para lograr autorreferencia en

de sucesor (el sucesor de x es el número natural inmediatamente siguiente a x).[28] De este modo, en **Q** se define 0 como el número que no es el sucesor de ningún número y con base en esto podríamos comenzar a generar la secuncia de los números naturales.

Tal secuencia es $0', 0'', 0''', ...,$ o 1, 2, 3,..., se genera a partir de la operación $'$ o $+0'$ (lo cual equivale simplemente a sumar 1). Con base en esto, se formulan las definiciones por inducción o definiciones recursivas de las funciones ($\phi(x)$) y los predicados ($P(x)$) del sistema: esto significa que primero se establece el valor de $\phi(0)$ o $P(0)$ y a partir de este valor se expresa $\phi(x')$ o $P(x')$ en términos de $\phi(x)$ y $P(x)$ y de esta manera podemos concluir que bajo estas circunstancias el valor de $\phi(x)$ o $P(x)$ se puede definir recursivamente para cualquier número natural x (Kleene, 1952, p. 217). Un ejemplo puede ayudar a explicar este concepto.

Para explicar cómo funciona la operación de suma (expresada como $\phi(x, y)$), sería necesario únicamente expresar cómo funciona la suma en el caso en que $y = 0$ y luego generalizar determinando cómo funciona para el sucesor de y. Esto nos arroja:

$$\phi(x, y) \begin{cases} x + 0 = x, \\ x + y' = (x + y)' \end{cases}$$

De tal manera que cualquier número x, si a este se le suma 0, el resultado es x, pero si a este se le suma el sucesor de y (y'), el resultado es igual a el sucesor de la suma de x y y. Para entender cómo estas ecuaciones nos permiten computar sumas, considérese sumar

---

lenguajes formales como la normalización, ver Smullyan (1957; 1994, Cap. 5, parte I).

[28] Los axiomas de **Q** son los siguientes:

1. $\neg(\mathbf{0} = x')$
2. $x' = y' \rightarrow x = y$
3. $x+\mathbf{0} = x$
4. $x+y' = (x+y)'$
5. $x\times\mathbf{0} = \mathbf{0}$
6. $x\times y' = (x\times y)+x$ (Raatikainen, 2018, sec. 1.2.1)

Nótese que es meramente la aritmética de Peano sin el esquema de inducción.

1 (1”) y 2 (0”). Lo que nos dice la ecuación es que $\phi(1,2) = 0' + 0'' = (0' + 0')'$. Ahora bien, $(0'+0')' = ((0'+0)')'$ por el segundo caso de la ecuación, y $((0'+0)')' = ((0'+0)')'$, lo que nos indica que el resultado de la suma entre 1 y 2 es el sucesor del sucesor de la suma entre 1 y 0. Dado que $x + 0 = x$ por el primer caso, $\phi(1,2) = 0'''$, es decir, 3.

Con esta herramienta lo que se busca es aritmetizar la sintáxis del sistema, asignándole a cada expresión un número natural como se explicó en la sección anterior. Cuando se hace esto, se le llama número de Gödel al número que representa la expresión del sistema y numeral de Gödel al numeral que le corresponde al número, es decir, si el número es n, el numeral será 0 acompañado de n signos de sucesor: si el número es 3, el numeral es $0'''$. En lo que sigue, utilizo n para referirme al número de Gödel de alguna expresión y **n** para referirme a su numeral de Gödel. De esta manera es posible representar todas las expresiones del sistema como de números o numerales.

Con esto en mente, se pasa a aritmetizar todas las expresiones del sistema progresivamente (lo que prueba que son recursivas) y con esto les habremos asignado un nombre a cada expresión. Normalmente se empieza con las operaciones lógicas de negación, un conectivo lógico y un cuantificador.[29] De esta manera, por ejemplo la prueba de que la negación es una función recursiva consiste en definir una función $neg(x)$ tal que $neg(x) = n{*}x$, donde * es simplemente la concatenación de ambos términos, es decir, si n es 3 y x es 56, la concatenación de ambos es 356. Así, $neg(x)$ representa la negación de cualquier fórmula del sistema y los valores se definen recursivamente, es decir, partiendo de 0 y a partir de ahí se establecen el resto de valores. Lo mismo se hace con el resto de operaciones lógicas.

Lo siguiente sería probar que la diagonalización es una función recursiva, es decir, probar que existe una función recursiva, $diag$, tal que si a es el número de Gödel de una expresión A, por lo tanto $diag(a)$ es el número de Gödel de la diagonalización de A. Recuérdese que la diagonalización consiste en tomar una fórmula con una variable libre y aplicarle la fórmula al nombre de la fórmula (la diagonalización de $B(x)$ sería $\exists x(x = \ulcorner B(x) \urcorner \wedge B(x))$, que es equivalente a $B(\ulcorner B(x) \urcorner)$).

Así como la función $neg(x)$ era $n * x$, hace falta especificar lo mismo de la diagonalización de manera que podamos expresar la diagonalización de cualquier expresión del sistema. La función de diagonalización hace uso de tres expresiones que no se encontraban

[29] Prueba: Proposición 15.1 de Boolos, Burgess y Jeffreys (Boolos y col., 1974/2007).

definidas previamente en el sistema: la cuantificación existencial ($\exists$), la conjunción ($\wedge$) y la asignación de un numero de Gödel.

Las tres operaciones son recursivas, la cuantificación existencial se representa como $exquant(v, x) = e * v * x$ donde e es el número de Gödel del signo $\exists$, v es el número de la variable y x la fórmula sobre la cual se cuantifica; la conjunción se representa como $conj(x * c * y)$ donde x es el número del primer término, c el de la coma y y el del segundo término; y la asignación de un numero de Gödel es $num(x)$.[30] La función de diagonalización es

$$diag(y) = exquant(v, conj(i * l * v * c * num(y) * r, y))$$

Donde v es el número de la variable, i,l, r y c son el signo igual a ($=$), el paréntesis izquierdo, el paréntesis derecho y la coma respectivamente.

Lo único que nos hace falta para poder formular el lema de diagonalización es poder representar esta función en el sistema **Q**. Es decir, hasta el momento en el sistema solo era posible expresar sintácticamente una serie de fórmulas bien formadas mediante números de Gödel, pero a pesar de este procedimiento no se podía determinar si una determinada función pertenecía al sistema, si hay una función dentro del sistema que represente la operación que se aritmetizó.

Una función recursiva f es representable en el sistema si hay una fórmula ψ_f tal que para cualquier a y b tal que sean números naturales, si $f(a) = b$ entonces $\forall y(\psi_f(\mathbf{a}, y) \leftrightarrow y = \mathbf{b})$ es la fórmula que representa $f(a) = b$ en el sistema y esta no solo es verdad relativo a este, sino también demostrable. Esto significa que todos los valores que puede tomar la función son demostrables en el sistema a partir de la aplicación de la función. Por ejemplo, si la función de suma ($f(a, b) = c$), debe existir una fórmula $\forall y(\psi_f(\mathbf{a},\ \mathbf{b}, y) \leftrightarrow y = \mathbf{c})$ que le asigna a cualquier a y b un único valor a partir de la aplicación de la función, esta fórmula es $\forall x \forall y \exists ! z(x+y = z)$.

A partir de los resultados presentados hasta aquí existen diversidad de procedimientos para probar que todas las funciones recursivas son representables en el sistema. Una

[30] La prueba de la recursividad de estas funciones está contenida en la proposición 15.1 y corolario 15.6 de Boolos, Burgess y Jeffreys (1974/2007).

prueba detallada de esto se encuentra en las proposiciones 16.1-16.16 de Boolos, Burgess y Jeffreys (1974/2007). Este resultado implica que los números de Gödel de todas las funciones recursivas del sistema son representables en $\mathbf{Q}$. Por lo tanto, la diagonalización (que es una función recursiva de $\mathbf{Q}$) es representable en el sistema. Así, $Diag(x, y)$ será la fórmula que represente la función $diag$, de manera que para cualquier m y n, si $diag(m) = n$, entonces $\forall y(Diag(\mathbf{m}, y) \leftrightarrow y = \mathbf{n})$ es la fórmula que representa la función de diagonalización en $\mathbf{Q}$.

Con esto en mente se puede formular el lema de diagonalización:

Lema 1. ***(Lema de Diagonalización o Teorema de Puntos Fijos)*** *Sea $B(x)$ una fórmula de $\mathbf{Q}$ con x como variable libre. Entonces hay una oración G tal que en $\mathbf{Q}$ se obtiene $G \leftrightarrow B(\ulcorner G \urcorner)$.*

Con base en este lema es posible formular oraciones autorreferenciales en aritmética. En esencia lo que indica es que para cualquier predicado del sistema existe por lo menos una expresión que dice de sí misma que tiene esa propiedad. Al ser este el resultado metamatemático más importante del capítulo, vale la pena detenerse en la prueba del lema. Para esto sigo a Boolos, Burgess y Jeffreys (1974/2007, 17.1 Lemma).

Prueba. Sea $A(x)$ la fórmula $\exists y(Diag(x, y) \wedge B(y))$. Sea a el número de Gödel de $A(x)$ y $\mathbf{a}$ su numeral. Sea G la oración $\exists x(x = \mathbf{a} \wedge \exists y(Diag(x, y) \wedge B(y)))$.

Por lo tanto, G es la fórmula $\exists x(x = \mathbf{a} \wedge A(x))$ (la diagonalización de $A(x)$) y es lógicamente equivalente a $A(\mathbf{a})$ y a $\exists y(Diag(\mathbf{a}, y) \wedge B(y))$. Por lo tanto,

$$G \leftrightarrow \exists y(Diag(\mathbf{a}, y) \wedge B(y))$$

Sea g el número de Gödel de G y $\mathbf{g}$ su numeral de Gödel. Dado que G es la diagonalización de $A(x)$, $diag(a) = \mathrm{g}$. Dado que es una función recursiva, hay una fórmula que la representa en el sistema (es decir, la diagonalización de la fórmula que es la diagonalización de otra es representable en el sistema). La fórmula que representa tal función es

$$\forall y(Diag(\mathbf{a}, y) \leftrightarrow y = \mathbf{g})$$

Dado que $\mathbf{g}$ es la diagonalización de $\mathbf{a}$, de esto se sigue que $G \leftrightarrow \exists y(y = \mathbf{g} \wedge B(y))$. Dado que $\exists y(y = \mathbf{g} \wedge B(y))$ es equivalente a $B(\mathbf{g})$, se sigue que

$$G \leftrightarrow B(\mathbf{g})$$

O, en otras palabras, $G \leftrightarrow B(\ulcorner G \urcorner)$. □

Con base en este método, de cualquier predicado definible recursivamente y representable en el sistema, como 'Verdadero' o 'Demostrable', puede deducirse una oración que diga de sí misma que tiene esa propiedad (Gödel, 1931/1986; Carnap, 1934/2001; Montague, 1962), i.e. que hay una oración que es la diagonalización de $\exists y(Diag(x, y) \wedge B(y))$ (donde B es la propiedad en cuestión) y tal oración tiene la propiedad B, es decir, una oración σ cuya cadena de referencia comienza y termina en σ.

3.3.3 Obteniendo la paradoja al estilo de Priest

Priest tiene una manera curiosa de obtener la paradoja del mentiroso, porque no la deriva aplicando el lema de diagonalización al predicado verdad (V), sino al de pertenencia ($\in$).[31] Como se explicó en el capítulo 2, para Priest la paradoja del mentiroso es una de las paradojas que cumplen con el *Inclosure Schema*, un esquema que está compuesto por las condiciones de Existencia, Trascendencia y Clausura. Estas condiciones se representaban de la siguiente manera:

1. Dado un conjunto Ω, tal que $\Omega = \{y : \phi(y)\}$ y $\psi(\Omega)$ (Existencia)
2. Si $x \subseteq \Omega$ y $\psi(x)$:
   (a) $\delta(x) \notin x$ (Transcendencia)
   (b) $\delta(x) \in \Omega$ (Clausura)

Como se expuso en el capítulo anterior, Priest utiliza la diagonalización para formular una oración α, que equivale a $\langle \alpha \notin a \rangle$, donde α es el número de Gödel de a, tal que $\delta(x)$

[31] Es posible diagonalizar fórmulas con más de una variable libre (Priest, 1987/2006, p. 48), de hecho es posible hacerlo para cualquier número arbitrario de variables gracias a un procedimiento elaborado por Montague (Montague, 1962, citado en (Picollo, 2020a, nota 5)).

es una oración $\sigma(a)$ y $\sigma(a) = \alpha$. Para simplificar las cosas, $\delta(x)$ será simplemente σ y $\sigma = \langle \sigma \notin a \rangle$, donde σ sigue siendo el número de Gödel de a.[32]

Así, cuando Ω era el conjunto de todas las oraciones verdaderas ($\{y : Tr(y)\}$), $\delta(Tr)$ arrojaba una oración σ de la forma $\langle \sigma \notin a \rangle$, donde $a = \Omega$. Tal oración ($\sigma = \langle \ulcorner \sigma \urcorner \notin Tr \rangle$) constituiría, para Priest la paradoja del mentiroso en tanto esta afirmaba que, según él, *esa* oración (σ) no era verdadera. De nuevo, para Priest esto captura todas las versiones de la paradoja independiente de la forma en la que se formule, ya sea con indéxicos o con demostrativos (1995/2002, pp. 290-291).

En los términos del aparato lógico presentado en este capítulo, lo que hace Priest para la paradoja del mentiroso es aplicar el lema de diagonalización predicado de no-pertenencia: $\sigma \leftrightarrow \neg P(\ulcorner \alpha \urcorner, Tr)$, donde $\sigma = \ulcorner \alpha \urcorner$ y $P(x, y)$ es el predicado que representa la relación de pertenencia en $\mathbf{Q}$, para simplificar las cosas, abreviaré $P(x, Tr)$ como $\Gamma(x)$. Por lo tanto, $\neg\Gamma(\ulcorner a \urcorner)$ equivale a $\exists x(x = \ulcorner a \urcorner \wedge \neg\Gamma(x))$. Por lo tanto, cuando σ es $\delta(Tr)$ (es decir, $\sigma = \langle \sigma \notin Tr \rangle$) lo que según Priest dice σ es 'esta oración no pertenece al conjunto de oraciones verdaderas', v.gr. esta oración es falsa.

Así, tomando en cuenta todo el proceso llevado hasta aquí, lo que expresa una oración como $\neg\Gamma(\ulcorner a \urcorner)$ (que es equivalente a σ) es que $\exists x(x = \ulcorner a \urcorner \wedge \exists y(Diag(x, y) \wedge \neg\Gamma(y)))$.

3.4 ¿Y qué dice σ?

Al lograr formular la paradoja del mentiroso en lenguajes formales mediante la diagonalización nos damos cuenta que, si bien los puntos fijos tienen una estructura $\psi \leftrightarrow \phi(\ulcorner \psi \urcorner)$, estos refieren a sí mismos en tanto son lógicamente a equivalentes a una oración de la forma $\exists x(x = \ulcorner \psi \urcorner) \wedge \exists y(Diag(x, y) \wedge \phi(y))$, es decir, en tanto son puntos fijos de un predicado (Picollo, 2018, p. 585). Y esto nos permite hacer varios comentarios respecto a lo que la fórmula refiere y cómo lo hace.

El primero es que basado respecto la referencia mediante cuantificadores. Una vez aplicamos las herramientas de la sección 3.2 a la fórmula $\exists x(x = \ulcorner a \urcorner \wedge \exists y(Diag(x, y) \wedge$

[32] De nuevo, esta es una simplificación presentada por Priest mismo en (2012, p. 368).

$\neg\Gamma(y)))$ se tiene que una vez se obtiene su forma normal postnexa disyuntiva[33] podemos afirmar que esta q-refiere a dos cosas, a una fórmula representada por $\ulcorner a \urcorner$ y a una fórmula y que es la diagonalización x y tiene la propiedad $\neg\Gamma$.[34]

Ahora bien, con base a esto podemos establecer que la fórmula es autorreferente, pues sabemos que $\ulcorner a \urcorner$ y y son la misma fórmula. Entonces no es cierto que este tipo de oraciones predican cosas de sí mismas en el sentido que, por ejemplo, lo hacen oraciones del lenguaje natural como 'estoy corriendo'. Solo podemos establecer que fórmulas en el lenguaje formal son autorreferentes porque sabemos que las dos fórmulas son equivalentes. Con base a esto, se ha llegado a afirmar que lo que sucede es que este tipo de fórmulas refieren de manera indirecta a sí mismas (Berto, 2009, p. 92), pero creo esto es poco preciso dada la definición de referencia directa ofrecida en la sección 3.2 (una fórmula ϕ refiere directamente a una fórmula ψ sii o m-refiere a ψ o q-refiere a ψ). En los casos obtenidos por diagonalización hay referencia directa, pues la oración q-refiere a sí misma, lo que pasa es que lo hace porque nosotros sabemos que $\ulcorner a \urcorner$ y y son la misma fórmula.

Cómo debe interpretarse la autorreferencia mediante diagonalización a la luz del lenguaje natural es un asunto poco estudiado y respecto del cual no hay consenso todavía. A la luz de lo expuesto en este capítulo me siento afín con un comentario de Milne al respecto, según el cual oraciones obtenidas mediante diagonalización refieren a sí mismas es similar a cuando alguien, sabiendo que es desorganizado y *queriendo decir* de sí mismo que es desorganizado, dice 'las personas que usan este apartamento', sabiendo que vive solo (Milne, 2007, p. 210).

33 La cual es $\neg\forall x(\neg(x = \ulcorner a \urcorner) \vee \forall y(Diag(x, y) \vee \Gamma(y)))$.

34 v.gr. 'no todo x' y 'para cualquier y', dada la forma normal postnexa disyuntiva.

Capítulo 4

Las paradojas del mentiroso en lenguaje natural: herramientas para su análisis

"*Is it really a belief about herself in a strict sense? The belief 'the last person who passed by lost the script' is only accidentally about herself, and she needn't know that it's about herself. If she thinks 'I lost that script', then it's not accidental – and only the latter belief prompts her to pick it up again.*"

Sophia, en el diálogo *The Root of Evil*

(Halbach, 2016, p. 159)

El propósito de este capítulo es presentar un aparato teórico con base en el cual se pueda analizar de manera satisfactoria las formulaciones de la paradoja del mentiroso en lenguaje natural. Con este espero poder demostrar no solo que estas difieren de la formulación de la paradoja en lenguaje formal, sino que las formulaciones en lenguaje natural difieren entre sí. Las formulaciones en lenguaje natural son, según Priest (Priest, 1995/2002, pp. 142-144):

- 'Estoy mintiendo' y
- 'Esta oración es falsa'.

La teoría que se presenta en este capítulo es la semántica de doble indexación de David

Kaplan, perfeccionada posteriormente por John Perry. De acuerdo con esta, existirían diferencias entre las formulaciones dada la forma en la que se logra la autorreferencia.

4.1 De Frege a Kaplan

El lenguaje natural tiene varias herramientas que nos permiten referirnos a cosas en el mundo. Los nombres propios son un ejemplo de esto. Podemos usar 'Jaakko Hintikka' para referirnos a un lógico y filósofo finlandés. Sin embargo, los nombres propios no son la única opción. También existen un tipo de expresiones llamadas deícticos, expresiones como 'yo', 'hoy', 'mañana', 'esto(a)', 'ahí', entre otras, cuya referencia (1) depende del contexto en el que sean usadas y (2) puede cambiar con un cambio de contexto.[35] Por ejemplo, 'mañana' utilizado el 30 de julio de 2019 se refiere al 31 de julio de 2019; pero utilizado el 12 de agosto de 2018 se refiere al 13 de agosto de 2018.

Aunque el significado de estas expresiones puede variar dependiendo del contexto, hay un nivel del significado de estas que es estable. Continuando con el ejemplo, a pesar de que 'mañana' refiere a días distintos según el día en el que se use, todos los usos deícticos de la expresión 'mañana' hacen referencia al día siguiente, y así con el resto de deícticos: 'yo' refiere a la persona que la emita, 'hoy' refiere al día en el que la palabra es emitida, 'esto(a)' refiere a un objeto particular, 'ahí' refiere a un lugar particular.

Las formulaciones de la paradoja del mentiroso en el lenguaje natural que se identificaron hacen uso de expresiones deícticas: 'estoy mintiendo' hace uso del 'yo';[36] 'esta

[35] En el mundo anglosajón estas expresiones son llamadas *indexicals* (Kaplan, 1989b, p. 490), *deictics* (Radulescu, 2018, p. 3173), *indicators* (Castañeda, 1989, p. 206) y *token-reflexive* (Reichenbach, 1947, p. 284). En el mundo iberoamericano han recibo los nombres *indéxicos* (Ezcurdia, 2014, p. 5) o *deíticos* (Gutiérrez Rodríguez, 2010, sec. 17). Para poder aprovechar la distinción indéxico-demostrativo, prefiero utilizar 'deíctico' para evitar confusiones.

[36] Aprovecho este espacio para hacer una aclaración sobre algunas expresiones que pueden sonar redundantes: en español, a diferencia del inglés, el verbo de la oración suele ser suficiente para determinar el pronombre de la misma, por lo que oraciones como en 'estoy mintiendo' se entiende que el pronombre de la oración es 'yo' dado que la conjugación 'estoy' es exclusiva de la primera persona singular. Así, es común que se omita el pronombre de la oración a menos que tenga un uso pleonástico, es decir, para dar énfasis o para evitar ambigüedad (Luján, 1999, § 20.1).

oración es falsa' requiere de 'esta'. De esta manera, para poder brindar un análisis satisfactorio de ambas formulaciones de la paradoja se requiere de una semántica que pueda conciliar el aspecto variable y el aspecto estable de los deícticos.

Una primer candidata para esto es la semántica que Gottlob Frege desarrolla en *Sobre sentido y referencia* (SuB, por sus siglas en alemán) (1892/2016) y ampliada en *El pensamiento: una investigación lógica* (DG, por sus siglas en alemán) (1918/2016). Según una posible interpretación, para Frege, hay dos niveles semánticos, el sentido (*Sinn*) y la referencia (*Bedeutung*). El sentido, por un lado, es el concepto o proposición que se expresa mediante cualquier expresión del lenguaje (Frege llama signos o nombres a estos últimos (1892/2016, p. 251)); por otro lado, la referencia es lo que se denota mediante el signo, es aquello de lo que el sentido es un concepto (Frege, 1892/2016, p. 255).

A manera de ejemplo, la descripción 'el séptimo ganador del premio Rolf Shock', en la teoría fregeana, expresa la proposición tal que 'existe solo un x tal que ese x es el séptimo ganador del premio Rolf Shock' y denota al objeto del mundo que cumple con esa propiedad, a saber, a Jaakko Hintikka. Como se ve en este ejemplo, el sentido es una manera de presentar una referencia, el sentido de un signo proporciona un criterio para identificar el objeto denotado. Pero puede que existan varios criterios de identificación para la misma referencia. Continuando con el ejemplo, para referirnos a Jaakko Hintikka pudimos haber usado la expresión 'el autor del libro *Knowedge and Belief*' y habríamos denotado a Jaakko Hintikka igualmente.[37]

Así, en la teoría fregeana, a un sentido le corresponde una referencia, pero a una referencia no le corresponde solamente un sentido (Frege, 1892/2016, p. 251). Esto permite que en una expresión significativa cualquiera de sus términos pueda ser reemplazado por otro con la misma referencia y se conserve el valor de verdad de la proposición original (Frege, 1892/2016, pp. 258-259): la oración 'Jaakko Hintikka fue un filósofo finlandés'

---

[37] Esta es una interpretación de la semántica fregeana en la cual el sentido de una expresión es entendida como una descripción del objeto que refiere y ha sido desarrollada por Bernard Linsky (1983, p. 10) y Edward Zalta (2019, sec. 3.2). Sin embargo, esta no es la única interpretación posible de la semántica fregeana, otras aproximaciones se han dado en términos de la relación que tiene el sentido con el valor de verdad de una oración (Dummett, 1973, caps. 1, 5, 17), en términos del sentido siendo análogo a una ruta que lleva a la referencia (Taylor, 1998, p. 7) o en términos del sentido como una manera de pensar la referencia (Evans, 1982, pp. 19-22).

sigue siendo verdad aunque se transformara y quedara 'el autor del libro *Knowedge and Belief* fue un filósofo finlandés'.

La teoría desarrollada en SuB es complementada por unos breves apuntes que Frege hace en DG y tienen que ver con la naturaleza del signo: hay ocasiones en las que el signo, que puede fijarse de manera escrita o en una grabación, no basta para expresar un concepto, pues, por ejemplo, hace falta determinar el momento de la expresión para que esta capte adecuadamente un pensamiento (Frege, 1918/2016, p. 330): si encuentro una nota en mi bolsillo que dice "la universidad va a estar abierta mañana", pero no sé la fecha en la que la escribí, no tengo manera de saber cuándo va a estar abierta la universidad.

De esta manera, Frege profundiza la relación entre un signo y el sentido que expresa en DG. Según él, en estos casos no es el texto que se conserva por escrito la expresión completa de un pensamiento, sino que se necesita conocer "ciertas circunstancias que acompañan a la emisión y que son utilizadas en ella como un medio para la expresión de un pensamiento" (Frege, 1918/2016, p. 330). Para el caso de la nota en mi bolsillo, el signo que expresa el pensamiento es tanto la expresión lingüística como algunas circunstancias propias de la emisión de la expresión, como el día en que fue escrita.

A pesar de esto, se ha considerado que la semántica fregeana es insuficiente para dar cuenta del significado de oraciones con deícticos (Perry, 1993a, pp. 7-9; Ezcurdia, 2014, p. 6). Esto se debe a que si bien Frege, en DG, quiere explicar el rol del contexto de las emisiones en la determinación de la referencia, el hecho de que el sentido de una emisión determine por sí mismo la referencia de la misma es problemático para dar cuenta del aspecto semántico estable y del aspecto semántico variable de los deícticos.

En términos fregeanos, parece que es posible capturar el aspecto semántico estable de los deícticos mediante la noción de sentido, de manera que podría pensarse que el sentido de 'mañana' sería el mismo que el de 'el día inmediatamente posterior'. Pero esto no es así. Supongamos que el 5 de agosto de 2019 le pregunto a mi hermano "¿la universidad va a estar abierta el 7 de agosto?". Si me responde "la universidad va a estar abierta mañana", se refiere a que la universidad va a estar abierta el 6 de agosto, pero si responde "la universidad va a estar abierta el día inmediatamente posterior" quiso decir que la universidad va a estar abierta el 8 de agosto, por lo que 'mañana' y 'el día inmediatamente posterior' no significan lo mismo.[38]

Casos de este tipo fueron introducidos para afirmar que no puede usarse la semántica fregeana para dar cuenta de las particularidades del significado de los deícticos. Sin embargo, hacia 1971 David Kaplan comienza a desarrollar una nueva semántica que tomaba en cuenta el problema de la teoría del significado fregeana y presenta una forma de solucionarlo.

Sin embargo, la teoría de Kaplan no ha sido la única que ha surgido en respuesta a Frege. En este trabajo se escoge la semántica de Kaplan porque con base en esta puede demostrarse la diferencia entre ambas paradojas de manera poco controversial y porque ha tenido la acogida más favorable entre otras rivales (Braun, 2017, introducción; Schroeter, 2017, sec. 1.1; Ezcurdia, 2014, p. 7).

4.2 Kaplan: carácter estable, contenido variable

Para poder dar cuenta del significado variable y del aspecto estable de los deícticos, Kaplan introduce dos diferencias: la primera es entre contenido y carácter, la segunda es entre contexto y circunstancia. En este apartado se explica cómo Kaplan utiliza estos cuatro conceptos para construir una semántica que pueda ayudarnos a examinar las diferencias que existen entre las dos formulaciones de la paradoja del mentiroso en lenguaje natural.

La propuesta de Kaplan parte de la distinción carnapiana entre intensión y extensión, la cual es similar a la distinción fregeana entre sentido y referencia, pero difieren en aspectos fundamentales (Carnap, 1947, §28-30; Parsons, 2016, sec. 4.2). Para Carnap, la extensión de un predicado es la clase de cosas para las cuales el predicado aplica y su intensión es la propiedad a la que refiere (1947, §4); la extensión de una oración es su valor de verdad, mientras que su intensión es la proposición expresada por esa oración (1947, §6); y la extensión de una expresión individual (una constante o una descripción) es el individuo que se denota mientras que su intensión es algo que Carnap llama *concepto individual* (1947, §7-9).

[38] Inpirado en un ejemplo dado por Perry en (1993a).

Para Carnap, una descripción-de-estado (*state-description*), o simplemente descripción, es una clase de oraciones de un lenguaje que para cada oración atómica (una oración que es verdadera o falsa y que no puede ser dividida en oraciones más simples) contiene o bien a esa oración o a su negación. Una descripción representa un estado posible del universo respecto todas las propiedades y relaciones expresadas por predicados del lenguaje.[39] Así, las nociones de intensión y extensión pueden entenderse de la siguiente manera: la intensión es una función que determina la extensión de un designador bajo una descripción particular.

Dada esta noción de descripción, podemos suponer, por ejemplo que la oración 'S' de un lenguaje L es verdadera en algunas descripciones, pero '$\neg S$' lo es en otras; sin embargo, '$S \vee \neg S$' siempre será verdadera, dada la definición de descripción-de-estado. De esta manera, dos designadores son lógicamente equivalentes (y por ende tendrán la misma intensión) si son verdaderos bajo las mismas descripciones.

Para poder dar cuenta del aspecto variable y del aspecto estable de los deícticos, Kaplan extiende la teoría carnapeana. Esto lo lleva a cabo en tres trabajos, *Dthat* (1978), *Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals* (1989b) y *Afterthoughts* (1989a).

Kaplan primero actualiza la semántica carnapiana a la kripkeana, desarrollada unos años antes de la publicación de *Demonstratives* a la que Carnap no tuvo acceso en su momento pero que estuvo sorprendentemente cerca de desarrollar por él mismo (Hintikka, 1975). Así, por ejemplo, cambió la noción de descripción por la de mundo posible accesible (Kaplan, 1989b, ver especialmente pp. 496-497).[40]

La primera distinción que Kaplan introduce para dar cuenta de los aspectos semánticos de los deícticos es entre contenido y carácter. Para él, el contenido de una expresión es lo que se dice, el significado de la oración en las circunstancias en las que se dice (Kaplan, 1989b, pp. 500-505). Para Kaplan, el contenido es una función que determina

[39] Según Carnap, descripción-de-estado captura la noción leibniziana de mundo posible o la wittgensteineana de estado posible de cosas (Carnap, 1947, p. 9).

[40] Para más información sobre los cambios concretos que requiere la teoría carnapiana para integrar la semántica kripkeana, ver (Ballarin, 2017, sec. 3.2). Para una explicación de la noción de mundos posibles, ver (Lycan, 1994).

la extensión de una expresión dadas unas circunstancias particulares, esto no es más que una reformulación de la noción de intensión.[41]

Así, el contenido de una expresión 'Φ' estaría representado por la siguiente fórmula (Forbes, 2003, p. 104):

1. $Con[\phi] = (< \text{Circunstancias} > \rightarrow Ext[\phi])$

Según Kaplan, el resultado de evaluar una fórmula bien formada 'α' en una circunstancia determinada es la extensión de 'α' (1989b, p. 501). Esto es similar a como debía determinarse la descripción relevante para determinar la extensión del designador en la semántica de Carnap. La diferencia entre la circunstancias kaplaneanas y la descripción carnapeana es que para Kaplan las circunstancias son el momento t y el mundo posible o estado de cosas w en los que se emite la expresión, representado mediante el par $< t, w >$ (Kaplan, 1989b, p. 498).[42] Para Kaplan esto se debe a que para determinar el contenido de expresiones sin deícticos no hace falta especificar más allá del momento en que fue emitida la expresión y el mundo posible relevante.

Veamos esto con un ejemplo. Dado el estado actual de cosas, en este momento (5 de agosto de 2019) es cierto que 'La universidad va a estar abierta el 6 de agosto de 2019', lo cual puede expresarse como

2. Fu

Así, para determinar el contenido de (2) hace falta tener en cuenta las circunstancias en las que se emite. El momento es el 05 de agosto de 2019 (t_0) y el mundo posible es el estado actual de cosas (w_0). Una función para determinar el contenido de (2) se expresa de la siguiente manera:

3. $Con[Fu] = (< t_0, w_0 > \rightarrow Ext[Fj])$

[41] Por expresión aquí me refiero a cualquier emisión significativa dicah en unas circunstancias particulares.

[42] Recuérdese que en la semántica kripkeana los mundo posibles son simplemente índices o puntos de evaluación.

Es así como la extensión de (2), un valor de verdad, se determina dependiendo del momento y del estado de cosas, sin hacer referencia otros parámetros. Sin embargo, este aparato aún no es lo suficientemente poderoso como para dar cuenta del significado de oraciones con deícticos, dado que aún cuando se actualice el aparato carnapiano a la semántica de mundos posibles, no es posible explicar el doble aspecto semántico (variable y estable) de las expresiones que contienen deícticos. Esto se puede ver mediante el siguiente ejemplo.

Si el 5 de agosto alguien dice "la universidad va a estar abierta mañana", dado el momento y el estado de cosas, el contenido de esta oración sería "la universidad va a estar abierta el 6 de agosto". Si bien esto es cierto, esto no captura el aspecto estable del significado de los deícticos: dadas estas circunstancias 'mañana' significa '6 de agosto', pero 'mañana' y '6 de agosto' no significan lo mismo en todas las circunstancias, basta con que sea dicho otro día para que 'mañana' signifique otro día distinto al 6 de agosto.

Para resolver este problema, Kaplan introduce la noción de carácter. Consideremos la siguiente regla:

> 'Mañana' refiere al día siguiente al que se emite la expresión.

Esta regla no nos dice a cuál día se refiere (la fecha), pero nos dice a qué día se refiere (el día siguiente). Esto se puede ver con reglas de otros deícticos, como 'yo'. La regla para 'yo' es que refiere a la persona que emite la expresión, nos dice a qué refiere el deíctico, pero no a quién refiere; para eso tenemos el contenido, que determina la extensión.

Estas reglas determinan el contenido de una expresión en distintos contextos y Kaplan las llama el carácter (Kaplan, 1989b, p. 505). El contenido, por su lado, se encarga de determinar la extensión de una expresión en unas determinadas circunstancias.

Ahora bien, debe notarse que circunstancia y contexto no son lo mismo. Como se explicó, las circunstancias son el conjunto de índices que son necesarios para determinar la extensión de una expresión sin deícticos, a saber, un momento t y un mundo posible w. Cuando Kaplan introduce la noción de carácter incluye otros índices que una vez establecidos permiten determinar la extensión de la expresión en una circunstancia particular.

Estos índices adicionales son el momento t de la expresión, el mundo posible w en el cual se utiliza la expresión, la posición p en donde se encuentra el agente de la expresión y el agente a de la expresión, representados por la 4-tupla $< t, w, p, a >$.[43] Así, gracias al contexto y al carácter, podemos dar cuenta del aspecto estable y el aspecto variable de las expresiones con deícticos.

Por esta razón, Kaplan considera que el carácter es una función de contextos a contenidos (Kaplan, 1989b). Esto se puede representar de la siguiente manera (Forbes, 2003, p. 104):

4. $Car[\phi] = (< t, w, p, a > \rightarrow Con[\phi])$

Una semántica como la de Carnap se conoce como una semántica de indexación, pues la determinación de la extensión de las expresiones depende de un índice, como la descripción-de-estado. Cuando en una semántica la determinación de la extensión de las expresiones depende de dos índices se le conoce como una semántica de doble indexación (Braun, 2017, sec. 3.2; Ezcurdia, 2014). La semántica de Kaplan es precisamente una semántica de doble indexación al requerir de dos índices, el contexto y las circunstancias (Kaplan, 1989b, pp. 509-510).

Volvamos al ejemplo. Si el 6 de agosto la universidad va a estar abierta y el 5 de agosto de 2019 (t_0), para este mundo posible (w_0), estando en Bogotá (p_0), Pablo Rivas Robledo (a_0) dice "la universidad está abierta mañana" (Ψ), el carácter de esta expresión se determinaría de esta manera:

5. $Car[\Psi] = (< t_0, w_0, p_0, a_0 > \rightarrow Con[\Psi])$

Donde $Con[\Psi]$ sería:

6. $Con[\Psi] = (< t_0, w_0 > \rightarrow Ext[\Psi])$

43 ¿Por qué es necesario tener t y w dos veces (tanto en el contexto como en las circunstancias)? Normalmente el momento y el mundo posible en los cuales se evalúa una oración es el mismo que el contexto en el cual se expresa la oración, por lo que el mundo posible w del contexto es el mismo que el de las circunstancias. Pero esto no siempre es el caso, pues el valor de verdad de una expresión con deícticos puede depender de un operador modal alético, como $\Diamond$ Fu. En este caso el mundo posible del contexto de la expresión no es el mismo mundo posible de las circunstancias de evaluación del valor de verdad de la expresión.

Ahora bien, la palabra “mañana” refiere al día inmediatamente posterior en cualquier contexto, usada en el contexto $< t_0, w_0, p_0, a_0 >$ y en las circunstancias $< t_0, w_0 >$ significa el 6 de agosto de 2019. De manera que Ψ emitida en el contexto y las circunstancias mencionadas significa que la universidad va a estar abierta el día inmediatamente posterior al 5 de agosto de 2019, esto es, va a estar abierta el 6 de agosto de 2019.

4.2.1 Indéxicos y demostrativos

Con la semántica de doble indexación que presentamos en la sección 4.2 es posible explicar, en términos generales, cómo determinar los valores de verdad de algunas expresiones que contienen deícticos. A pesar de esto, debemos explicar cómo funciona esto para los dos tipos de deícticos que existen. Algunos deícticos solo requieren del contexto para saber a quién, qué lugar o qué día se refieren: por ejemplo, para expresiones que contengan deícticos como ‘yo’, ‘hoy’, ‘mañana’ o ‘ayer’ podemos determinar su referencia sin ir más allá del contexto de emisión. A estos deícticos se les denomina, siguiendo la terminología de Kaplan, indéxicos (Kaplan, 1989b, p. 490).

Por otro lado, existen algunas expresiones que, a pesar de ser deícticas, requieren más que del contexto para determinar su referente, como ‘este(a)’, ‘él’, ‘ella’, ‘ahí’ o ‘allá’, a estos deícticos se les denomina, siguiendo la terminología de Kaplan, demostrativos. Por ejemplo, si alguien afirma “él toma café” sin más, de un momento a otro, sin estar cerca de nadie, esta oración no tiene una referencia determinada. Por esta razón, hace falta más que el contexto para determinar la extensión de expresiones que contengan este tipo de deícticos.

Esta sección busca presentar una buena explicación que la teoría kaplaniana puede dar para dar cuenta del significado de expresiones demostrativas, lo cual es esencial para analizar la segunda formulación de la paradoja del mentiroso en lenguaje natural, “esta oración es falsa”. Este desafío cuenta con una dificultad adicional, y es que a lo largo de su trabajo Kaplan presentó 2 posibles formas en las cuales su semántica podía darle cabida a las expresiones demostrativas, las cuales son incompatibles entre sí. La última de estas versiones aparece en *Afterthoughts* (1989a, pp. 582-590), donde Kaplan corrige los errores de su primera teoría, pero no es desarrollada a profundidad por él.

Por lo tanto, a continuación se reconstruye la última versión de la teoría de Kaplan, la teoría de las intenciones directrices con unos comentarios adicionales. En aras de robustecer la teoría de las intenciones directrices, constantemente se acude al valioso texto de John Perry, *Directing intentions* (2009), en tanto (i) Perry se toma enserio la propuesta de Kaplan y procura complementarla y (ii) la teoría de las intenciones directrices resulta ser sorprendentemente útil para establecer las diferencias que existen entre las paradojas en lenguaje natural.

La teoría de las intenciones directrices

En el marco inicial, Kaplan le daba más importancia a la demostración que acompañaba la oración que a la intención con la que se dirigía la oración o la intención directriz (*directing intention*). La diferencia entre estos dos es la siguiente: mientras que la primera es una externalización con la cual se señala el objeto denotado por el demostrativo, la intención directriz es la acción mental de dirigir el pensamiento demostrativo hacia un objeto.

Pero supongamos que en una fiesta quiero referirme a una mujer que sujeta un vaso de martini y pregunto '¿quién es la mujer del martini?' Sin embargo, resulta que el vaso contiene agua y no un martini. En este caso, la demostración era 'la mujer con el martini', una descripción definida; la intención con la que dirigí la oración era mi deseo de referirme a esa mujer. En ese mismo caso, la demostración verbal refiere a una mujer con un martini que está en la misma sala que yo, pero la intención directriz refería a una mujer que sostenía una copa en la que usualmente se sirven martinis. Pero dado que no hay una mujer con un martini, mi demostración no refiere a algo, aunque mi intención directriz sí.[44] Precisamente por casos como estos, Kaplan corrige su teoría y pasa a darle más importancia a la intención directriz que a la demostración (Kaplan, 1989a, p. 582).

Para Kaplan, las dos características principales de las intenciones directrices son las siguientes:

1. Hacen parte del carácter de la oración y

[44] Ejemplo adaptado del propuesto por Keith Donellan en (1966, p. 287).

2. es el elemento semántico que permite diferenciar diferentes ocurrencias del mismo demostrativo, pues cada ocurrencia de un demostrativo debe ser provista de una intención directriz particular[45] (Kaplan, 1989a, pp. 587-589).

Así, de acuerdo con lo primero, 'él', 'ella', 'aquel', 'ese' o 'esos' se convierten en 'él[δ]', 'ella[δ]', 'aquel[δ]', 'ese[δ]' o 'esos[δ]', donde 'δ' representa la intención directriz. De acuerdo con lo segundo, pueden haber diferentes ocurrencias del mismo demostrativo en una misma oración y referir a objetos distintos.

De esta manera, el carácter de la oración "él toma café" cuando es usada por alguien que quiere referirse a un varón cercano que toma café se analiza como "él[δ] toma café". Aquí, el carácter de 'él[δ]' es el varón al que el agente o emisor de la oración se quiere referir en ese contexto. Esta oración puede venir acompañada de una demostración que ayude a los receptores de la oración a entender a quién se está refiriendo el emisor, pero basta con que el emisor se quiera referir a esa persona para que haya referencia a esa persona, esa es la ventaja de las intenciones directrices.

Así, en el caso de la mujer del martini, se podría decir que quien dice "¿quién es la mujer con el martini?" no se está refiriendo a una mujer con un martini, sino a la mujer que está viendo. Esto es posible, según Perry, si le podemos adjudicar al emisor la intención de referirse a esa mujer. Es decir, si el emisor planeaba con su intención directriz referirse a la mujer que estaba viendo (Perry, 2009, p. 190).

Para Perry, la clave está en ver las intenciones directrices como algo que normalmente está ligado a *lo que dice* una oración (o contenido locucionario), pero que puede estar desligado del mismo (Perry, 2009, pp. 192-193).[46] Esto, según Perry, es especialmente cierto en los casos en los que se referencia mediante demostrativos.

Así, para Perry, los casos anteriores se explican en tanto hay algo que se intenta y no se logra: en el caso de "¿quién es la mujer con el martini?" quien preguntaba quería referirse a una mujer de la fiesta que tenía un martini, pero no era el caso que lo tuviera, por lo que sale mal. El plan de quien pregunta depende de que en efecto la mujer a la que se está refiriendo tenga un martini, lo cual es falso. De manera que quien pregunta

[45] Para Kaplan, esto último no es cierto para oraciones anafóricas (1989a, pp. 588-589).

[46] Perry toma este término de la filosofía de J.L. Austin (Perry, 2009, p. 192).

hace lo que quería hacer (referirse a la mujer en la fiesta) pero no logra hacer lo que tenía la intención de hacer al referirse a la mujer como "la mujer del martini" (Perry, 2009, p. 192).

En consecuencia, una oración con demostrativos significativa siempre debe estar acompañada por una intención directriz que permita evaluar el carácter en un contexto y el contenido en unas circunstancias. El carácter de 'ella[δ]' refiere a la mujer con la cual el agente de la emisión $< a >$ tiene cierta conexión (muchas veces perceptual) en $< t, w, p >$, esto permite que el referente de 'ella[δ]' sea la mujer con la cual el agente tiene una conexión (Perry, 2009, p. 196).[47]

4.3 'Yo' vs. 'esta'

Este capítulo termina planteando dos diferencias relevantes que existen cuando se hace referencia por medios de los deícticos 'yo' y 'esta'. Estos dos deícticos aparecen en las formulaciones de las paradojas del mentiroso en lenguaje natural que se identificaron y será relevante en el capítulo siguiente. Recuérdese que las formulaciones son:

- 'Estoy mintiendo' y
- 'Esta oración es falsa'.

Por supuesto, la primera refiere a una oración dicha por un agente en un contexto y la segunda a una oración; 'yo' es un indexical y 'esta' es un demostrativo. Esta última diferencia, por más que obvia, pone de manifiesto una diferencia semántica que existe a la hora de referirse con indéxicos y con demostrativos.

Tal diferencia puede resumirse de la siguiente manera: al menos en el caso del indéxico 'yo', el contexto es suficiente para que el receptor determine el contenido de la oración (y por lo tanto su referente) en esas circunstancias de manera automática (Perry, 1997,

[47] Perry afirma que la conexión que tiene el agente con el referente no debe ser precisamente perceptual, pues puede ser mental o incluso podemos tener el mismo tipo de conexión con algo que solamente recordemos, esto lleva a que, para él, en 'ella[δ]' la letra 'δ' no es precisamente la intención directriz, sino el tipo de intención directriz, según el tipo de conexión que el agente tenga con el referente (Perry, 1993a, p. 196).

pp. 595-596; Kaplan, 1989b, p. 489); al menos en el caso del demostrativo 'esta' (y sus variantes lingüísticas, como 'este', 'estas', 'estos'), el contexto no siempre es suficiente para que los receptores determinen el contenido de la emisión en esas circunstancias, muchas veces hace falta un elemento adicional que ayuda a determinar el referente de la emisión (Perry, 1997, p. 196; Kaplan, 1989b, p. 490).[48]

Por ejemplo, si estoy viendo fotos y encuentro una foto en la que está toda mi familia, y en la que además yo salí muy mal, y le digo a mi hermano "qué mal estoy en esta foto" no hay duda de que me estoy refiriendo a mí. Pero si en la misma situación, queriéndome referir a mí mismo,[49] digo "qué mal salió esta persona en esta foto", puede que para mi hermano no quede claro a quién me quiero referir, si a él o a mí o a cualquier persona que sale en la foto. Esto, dado que si bien quería decir que yo era el que salía mal en la foto, no fui lo suficientemente claro. Para eliminar lo vaga que fue mi expresión, podría haberla acompañado de una demostración, pero como se intentó resaltar en la subsección anterior, esto es contingente.

Además de esta diferencia semántica existe otra diferencia importante cuando nos referimos a la misma cosa mediante el indéxico y el demostrativo ya referido. El indéxico 'yo' refiere únicamente al agente de la emisión, no puedo utilizar 'yo' para referirme a otra persona o cosa que no sea yo mismo.[50] Pero, por otro lado, con 'esta' me puedo referir a muchas cosas del mundo, incluído a mí mismo (por ejemplo, si digo 'esta persona es un abogado que está terminando la carrera de filosofía' mientras señalo una foto mía). Es decir, cualquier oración con que contenga 'yo' es autorreferente, pues se refiere a la persona que emite la oración; mientras que una oración que contenga 'esta' puede ser autorreferente por referirse a la persona que emite la oración o a la oración que se emite ('esta oración tiene cinco palabras'). Pero esto no significa que todas las oraciones que tengan 'esta' son autorreferentes, por ejemplo, "esta mujer tiene un martini".

48 Es imperioso leer esto último a la luz de la teoría de las intenciones directrices: para el emisor, se puede determinar el contenido en las circunstancias de emisión porque tiene una conexión con el objeto y puede tener una intención directriz hacia este; el elemento adicional aparece para que el receptor determine el contenido en los casos en los que no es claro a qué o quién se está refiriendo el emisor, como en el caso de las fotos.

49 Es decir, mi intención directriz está constituida por una conexión mental conmigo mismo.

50 Excluyo de esta regla el lenguaje indirecto, pues en este los deícticos tienen un uso anafórico.

El hecho que las oraciones con el indexical 'yo' tengan un ámbito de referencia tan limitado ha llevado a afirmar que este tipo de oraciones ocupan un lugar especial dentro de las teorías del significado, pues no son expresables en otros términos más que con ese indexical (Castañeda, 1966; Perry, 1993b; Lewis, 1979; Stalnaker, 1981). Perry, en concreto, lo explica de la siguiente manera: las oraciones que hacen uso de 'yo' tienen un ingrediente adicional que hace que la proposición expresada por la oración no pueda ser expresada por una oración que omita ese indexical (Perry, 1993b, p. 37). Por supuesto, la teoría de Perry solo es una teoría plausible, pero esta resulta tener una especial afinidad con la semántica de Kaplan.

Supongamos que estoy en un supermercado y noto que hay un camino de azúcar, como si alguien en su carrito tuviera un paquete de azúcar roto y se le estuviera regando, y pienso "a esta persona se le está regando el azúcar, voy a buscarla para avisarle", pues soy una persona que se preocupa porque el supermercado esté limpio. Así, empujo mi carrito alrededor de uno de los pasillos buscando a la persona que está regando azúcar para avisarle del desastre que está haciendo. Sin embargo, conforme voy siguiendo el camino, me doy cuenta que este se va haciendo cada vez más grueso, como si la persona que está regando el azúcar ya hubiera pasado por ahí. De repente me doy cuenta que "yo estoy regando el azúcar". Por lo que me detengo y acomodo el paquete que está regando el azúcar (ejemplo tomado de Perry (1993b, p. 33)).

En el ejemplo me refiero a mí mismo de dos maneras. La primera, mediante el demostrativo 'esta' ("a *esta* persona se le está regando el azúcar(...)"); la segunda, mediante el indéxico 'yo' ("*yo* estoy regando el azúcar"). Sin embargo, parece que la primera y la segunda manera de referirme a mí difieren de en un aspecto esencial, pues cuando me referí a mí mismo con 'esta' no me paré a acomodar el paquete de azúcar. Esto solo lo hice cuando me di cuenta que era yo quien estaba regando el azúcar, v.gr. cuando me referí a mí mismo con 'yo', antes de eso no sabía que la persona que estaba regando el azúcar y yo éramos la misma persona.

Casos como este han llevado a afirmar que en las expresiones acerca de uno mismo (actitudes *de se*, en palabras de Lewis (1979, p. 521)) los indexicables sean ineliminables (Castañeda, 1966, § III) o inexplicables en otros términos (Lewis, 1979, pp. 521, 527-8; Perry, 1993b, pp. 33-45). Esto, en tanto lo que se expresa en oraciones acerca de uno mismo con el indéxico 'yo' es inexpresable mediante otro tipo de expresiones, incluso otros deícticos.[51]

Exponer detalladamente por qué esto es cierto para cada tipo de expresión se sale de los límites de esta investigación, la forma particular en la que esta diferencia es relevante para nuestros propósitos es la imposibilidad de reemplazar en una oración el indéxico 'yo' por el demostrativo 'esta' tal como la concibe Perry (1993b). Según Perry, en oraciones como "yo estoy regando el azúcar" no se podría reemplazar el 'yo' por 'esta persona[δ]' en tanto no sería posible explicar por qué pasé de buscar a la persona que estaba regando el azúcar a acomodar el azúcar que se me estaba regando (1993b, pp. 33, 47-8).

Esto se debe a que, en el ejemplo del supermercado, de la expresión "a esta persona se le está regando el azúcar" no se sigue "yo estoy regando el azúcar". Esto, pues es posible que no sepa que la estoy regando, si bien en el primer caso me estoy refiriendo a mí, no sé que me estoy refiriendo a mí. Por el contrario, si desde el principio afirmo que "yo estoy regando el azúcar" no existe duda de que sé que me estoy refiriendo a mí, pues el carácter de este indéxico hace que refiera al agente de la emisión en todos los contextos de evaluación, tal como se vio en la subsección 4.2.1. En el primer caso, del que no sepa que soy yo quien está regando el azúcar se sigue que no me detengo a acomodar el paquete roto, y por lo tanto no sería posible explicar un cambio en mi conducta.

Así, todas la veces en las que en este trabajo haya usado la palabra 'yo' he estado diciendo cosas sobre mí. En el ejemplo de la mujer del martini era yo quien estaba en el salón y preguntaba por ella; en el ejemplo de las fotos era yo quien estaba con mi hermano viendo fotos; en el ejemplo del supermercado, era yo quien regaba el azúcar.[52]

---

[51] Vale la pena resaltar que esta línea de pensamiento no es universalmente aceptada y ha sido criticada argumentado que no hay tal cosa como un indéxico esencial o actitudes *de se* sino que son una mera ilusión (Cappelen & Dever, 2009) o un mito (Magidor, 2015). Quienes defienden que existen actitudes *de se* y que estas son irreducibles a otros tipo de actitudes son llamados *excepcionalistas de se*, mientras que quienes niegan alguna parte de esa conjunción conocidos como *escépticos de se* (Ninan, 2016).

[52] He evitado intencionalmente hablar de mí en términos de Pablo Rivas Robledo. Precisamente porque puede que al referirme a mí por mi nombre propio haya casos en los que no sepa que estoy hablando de mí (i.e. cuando se me olvida mi nombre y hablo sobre Pablo Rivas Robledo), algo que se evita al usar 'yo'.

Esto también es cierto respecto de cada vez que haya escrito con indexical ‘estoy mintiendo’. Esta siempre ha sido mi paradoja, dado que cada vez que la escribo, estoy en un contexto tal que el redactor de esa oración soy yo. Por esto, cada vez que he escrito la paradoja he dicho algo acerca de mí, que estoy mintiendo.

Capítulo 5

Paradojas del mentiroso

“*In the literature people tend to describe sentences obtained via Godel’s diagonal construction as sentences saying: ‘I am X’. But, as I just said, that strikes me as unjustified*”

Sophia, en el diálogo *The Root of Evil*

(Halbach, 2016, p. 162)

En el segundo capítulo de este trabajo se explicaron las razones por las cuales Priest piensa que todas las paradojas de la autorreferencia tienen la misma estructura (*Inclosure Schema*) y cómo esto lo lleva a afirmar que tienen una solución uniforme. Además de esto, en ese capítulo también se remarcó cómo, a pesar de que haya paradojas que tengan la misma estructura, Priest diferencia paradojas de la autorreferencia en tanto difieran en la forma de lograr autorreferencia y el tipo de objeto al que refieren. Por último, allí se explicó en detalle el tratamiento que le da Priest a la paradoja del mentiroso: si bien reconoce que hay varias formulaciones de la paradoja (tanto en lenguaje formal como en el natural), piensa que todas son reducibles a una única expresión de la paradoja en lenguaje formal, σ, en tanto para la paradoja del mentiroso las diferencias entre las formas de lograr autorreferencia son irrelevantes.

En el tercer capítulo se examinó con aún más detalle la formulación de la paradoja en lenguaje formal, encontrando que esta es un existencial afirmativo, un punto fijo del predicado de no pertenencia al conjunto de oraciones verdaderas obtenido a partir de la diagonalización. Por último, teniendo en cuenta que las dos formulaciones de la paradoja en lenguaje natural ('estoy mintiendo' y 'esta oración es falsa') hacen uso de deícticos, el cuarto capítulo se encargó de explicar cómo funcionaban tales expresiones.

Es hora de comparar las tres formulaciones y determinar si estas son distintas. Para ello primero utilizo el aparato técnico presentado en los dos capítulos anteriores y examino cada una de las formulaciones de la paradoja y qué hace paradójicas a estas expresiones, es decir, el tipo de objeto a las que estas expresiones hacen referencia y cómo las paradojas logran autorreferencia. Utilizo esto para posteriormente argumentar que con σ es imposible expresar alguna de las dos formulaciones de la paradoja en lenguaje natural. Por último examino algunas consecuencias de este hallazgo para el Principio de Solución Uniforme, el *Inclosure Schema* y el dialeteísmo.

5.1 Lo paradójico del asunto

La primera formulación de la paradoja del mentiroso a examinar es 'estoy mintiendo', en la cual el indexical 'yo' es determinante a la hora de crear la paradoja. Con base en el aparato expuesto en el cuarto capítulo, podemos ver que la referencia de esta oración cambia de acuerdo con las circunstancias: dado un agente particular, la oración 'estoy mintiendo' hace referencia a una oración dicha en un contexto particular por ese agente (que en este caso es la oración misma que se emitió).[53]

De este modo, la referencia de la oración cambia cada vez que hay un cambio de agente: si Jaakko Hintikka dice 'estoy mintiendo' está haciendo referencia a algo distinto a cuando yo digo 'estoy mintiendo'. En el primer caso la oración hace referencia a la oración que dice Jaakko Hintikka: $Car[\text{Estoy mintiendo}] = (< t, w, p, a_{Jaakko} > \rightarrow Con[\text{Estoy mintiendo}])$; pero en el segundo hace referencia a la oración que yo digo: $Car[\text{Estoy mintiendo}] = (< t, w, p, a_{Pablo} > \rightarrow Con[\text{Estoy mintiendo}])$.

No obstante, una diferencia a este nivel no hace que la paradoja dicha por mí y la paradoja dicha por Hintikka sean paradojas distintas, sino las hace diferentes instancias de la misma paradoja, pues las expresiones tienen el mismo carácter. Las diferencias aparecen

[53] Es necesario tener en cuenta que en este trabajo (1) se limitó el sentido de 'mentir' a simplemente emitir intencionalmente oraciones falsas y (2) que se está presuponiendo el hecho que las oraciones son portadoras de la verdad.

en el contexto de emisión. Nótese que en los ejemplos del párrafo anterior en ambos casos se intentaba determinar el carácter de 'estoy mintiendo' (Car[Estoy mintiendo]). Así, en la semántica de Kaplan-Perry, todas las emisiones de 'estoy mintiendo' tendrían el mismo carácter, pero el contenido puede variar cuando varía el contexto de emisión. En estos casos el contexto es relevante

De hecho, dado que el carácter es un elemento del significado que es estable a pesar del cambio de circunstancias o de contexto, todas las emisiones de 'estoy mintiendo' tienen el mismo carácter; pero estas difieren al nivel de su contenido cuando son emitidas por agentes distintos, pues el contenido de la oración está determinado por las circunstancias de emisión (en concreto, el agente emisor). Por lo tanto, lo que distingue de manera no trivial una instancia 'estoy mintiendo' otra es el agente de la emisión: dos agentes no pueden emitir la misma instancia de la paradoja.

La segunda formulación de la paradoja en lenguaje natural cuenta con un demostrativo: '*esta* oración es falsa'. Aquí es relevante traer la teoría de las intenciones directrices. Así, la paradoja sería 'esta[δ] oración es falsa', donde δ es la intención directriz que tiene el emisor de la oración tal que está dirigida a la oración que está emitiendo.

Dado que la intención directriz hace parte del carácter de la emisión, es importante que la intención directriz esté volcada a la oración que se está emitiendo, de lo contrario no habría paradoja. Por ejemplo, si cuando leo un libro de historia me doy cuenta que una de las aseveraciones de la autora es falsa, puedo decir 'esta[δ] oración es falsa' donde δ dirige a la oración del libro de historia y no hay paradoja.

Esta versión de la paradoja es sobre oraciones (no necesariamente oraciones emitadas por el agente emisor) y los cambios de agente solo son importantes en la medida en la que también cambie la intención, contrario a lo que sucedía con las instancias de la primera formulación de la paradoja. Por el contrario, esta formulación requiere de un demostrativo y de una intención directriz particular. De modo que también hay distintas instancias de la paradoja cada vez que la intención directriz del emisor está volcada hacia una oración distinta.

De este modo, las dos formulaciones de la paradoja en lenguaje natural que identifica Priest en *Beyond the Limits of Thought* difieren en dos aspectos fundamentales: (1) variantes de la primera versión de la paradoja ('estoy mintiendo) difieren en tanto sus circunstancias son distintas, en el caso de las variantes de la segunda formulación ('esta oración es falsa'), difieren en tanto su carácter es distinto (recuérdese que las intenciones directrices hacen parte del carácter); (2) mientras que dos personas no pueden instanciar la primera formulación de la paradoja de la misma manera, dos personas sí pueden hacerlo con respecto la segunda formulación de la paradoja, pues se trata de hacer referencia a oraciones con demostrativos: una persona puede decir 'esta[δ] oración es falsa' y otra persona puede decir 'esa[δ] oración es falsa'. Mientras que δ en ambos casos esté dirigida a la misma oración, ambas personas habrán instanciado la misma paradoja dado que el carácter no habrá cambiado.

Si la primera versión de la paradoja en lenguajes naturales es sobre oraciones de un agente particular y la segunda versión es sobre oraciones, la versión de la paradoja en lenguaje formal, por su lado, q-refiere a números de Gödel que son el nombre de fórmulas. En efecto, la paradoja en lenguaje formal es una oración σ tal que

$$\sigma = \exists x(x = \ulcorner a \urcorner \wedge \exists y(Diag(x, y) \wedge \neg\Gamma(y)))$$

La expresión del lado derecho de la equivalencia, que es la expresión autorreferencial, está diciendo algo de x y en este caso es el número de Gödel de σ. Así, la paradoja radica en que hay un número de Gödel (y) que es la diagonalización de $\ulcorner\sigma\urcorner$ y que y no pertence al conjunto de oraciones verdaderas. En este caso, σ es la fórmula representada por ese número de Gödel.

Tenemos pues un enunciado acerca de una oración del agente que la emite que es paradójico gracias al uso autorreferencial de un indéxico; un segundo enunciado acerca de oraciones que es paradójico gracias al uso autorreferencial de un demostrativo y una intención directriz; y un tercer enunciado que es paradójico gracias al uso autorreferencial de la diagonalización.[54] Llegó la hora de preguntarnos si el tercero de estos puede expresar los otros dos.

[54] Aquí el "gracias a" es usado para expresar que, por ejemplo, el uso del indéxico es una condición para lograr la paradoja.

5.2 El poder de reducibilidad de σ

En esta sección me concentro en explicar cómo las versiones de la paradoja en lenguaje natural no son expresables en los términos de la formulación de la paradoja en lenguaje formal, es decir, en cómo no se puede expresar las versiones de la paradoja en lenguaje natural mediante un punto fijo del predicado de no pertenencia al conjunto de oraciones verdaderas. Para hacerlo, exploro los límites del sistema **Q** y de la función de diagonalización, concluyendo que las limitaciones propias del sistema nos impiden hacer referencia a algo más que números y que la diagonalización es un mecanismo limitado de autorreferencia con el cual no es posible crear diferentes instancias de las paradojas.[55]

Así, dado que el proyecto en cuestión es revisar el tratamiento que Priest les da a las formulaciones de la paradoja, no hace falta establecer si el proceso inverso es posible (si se puede reducir σ o bien a un enunciado autorreferente mediante el indéxico 'yo' o bien a un enunciado autorreferente mediante el demostrativo 'esta'). De nuevo, Priest lo que afirma es que en tanto la forma precisa en la que se logra autorreferencia no es relevante, las expresiones 'estoy mintiendo' y 'esta oración es falsa' se reducen a σ, son la misma paradoja (Priest, 1995/2002, p. 280). En lo que sigue muestro que sí es relevante.

5.2.1 ¿Podemos expresar 'estoy mintiendo' mediante σ?

Las dos diferencias clave entre la versión de la paradoja con indéxico y la formulación en lenguaje formal es que la primera utiliza el indéxico y es sobre una oración emitida en un contexto por un agente, mientras que la segunda utiliza la diagonalización y q-refiere un número que representa una oración mediante un cuantificador existencial afirmativo. Así, si bien ambas son respecto fórmulas u oraciones, la diferencia crucial es que en el caso de la paradoja en lenguaje natural es necesario apelar a elementos contextuales para poder darle sentido a lo que se dice, mientras que en la paradoja en lenguaje formal ni puedo hacerlo ni necesito hacerlo. Para mayor claridad recuérdese que las formulaciones son:

1. Estoy mintiendo; y

[55] Para más información sobre el sistema **Q**, ver la subsección 3.3.2 de este trabajo.

2. $\exists x(x = \ulcorner a \urcorner \wedge \exists y(Diag(x, y) \wedge \neg\Gamma(y)))$.

Dos preguntas surgen de estas formulaciones de la paradoja: ¿puede la oración 2 referir a oraciones que yo emito? y ¿puede la diagonalización cumplir la misma función que el indéxico 'yo' (en tanto generador de autorreferencia)? Teniendo en cuenta lo expuesto en los capítulos 3 y 4 de este trabajo, en esta sección defiendo que la respuesta a lo primero es que no, esto se debe a la naturaleza misma del sistema **Q**, para el cual es imposible hacer m-referencia o q-referencia a objetos por fuera del sistema; la respuesta a lo segundo es que no se puede debido a las grandes diferencias que existen entre el 'yo' y la diagonalización como mecanismos de autorreferencia, las cuales explico en esta sección.

El sistema **Q** es un sistema que está pensado para demostrar una serie de enunciados matemáticos sobre los números naturales. Así, la diagonalización busca que el sistema hable de sí mismo utilizando los mismos términos en los que está formulado el sistema: usa los números naturales como nombres de expresiones del sistema y así garantiza que los nombres mismos hagan parte del sistema.

Sin embargo, el mecanismo de referencia sistema **Q** (numeración gödeliana) no permite hablar de cosas por fuera del sistema, como agentes o las oraciones emitidas por estos, porque precisamente el mecanismo no está pensado para nombrar cosas más allá de **Q**. Así, por ejemplo, no podría hablar sobre mí utilizando el sistema **Q** precisamente porque no soy un objeto sobre el cual el sistema **Q** pueda hablar. De esta manera en **Q** no puedo decir cosas como 'estoy corriendo', 'estoy hablando' o 'estoy mintiendo'.

Enfoquémonos en la segunda diferencia, el medio para generar autorreferencia. En este caso teníamos que la paradoja en lenguaje formal utiliza la diagonalización y el cuantificador, mientras que su contraparte en lenguaje natural, en este caso, utiliza el indéxico yo.

Respecto a esta diferencia en particular, creo que no puede haber un caso en donde se logre autorreferencia con el indéxico 'yo' que también pueda ser expresado de la misma manera diagonalizando. Para dar argumentos en pro de esta posición es necesario volver a varios puntos de los capítulos 3 y 4 de este trabajo.

Uno de los aspectos del indéxico 'yo' más relevantes era su naturaleza deíctica. Como todo deíctico, un aspecto del significado de 'yo' es variable, su contenido. Pero hay otro aspecto que es estable, y es la regla que determina el significado de la expresión en el contexto de emisión, que Kaplan llama carácter (Kaplan, 1989b, p. 505).

Como vimos, cuando en una oración el sujeto de la oración se expresa con 'yo', el carácter de este indéxico hace que la oración sea sobre el agente que emite o escribe la oración en ese contexto. Con el indéxico 'yo', sucede que dos oraciones pueden tener el mismo carácter pero tener distinto contenido si varía el agente de emisión: si Jakko Hintikka dice 'estoy mintiendo', está diciendo algo distinto a cuando yo digo 'estoy mintiendo'. Dado que no hay cambio en el carácter, hay un nivel del significado en el cual ambas oraciones coinciden (el carácter), pero hay uno en el que difieren (el contenido).

De esta manera, cuando digo 'estoy mintiendo' estoy diciendo algo acerca de una oración que estoy diciendo, a saber, que es falsa. Pero si la escribe cualquier otro agente, la paradoja recae sobre una acción de ese agente, de modo que no es posible que dos personas instancien exactamente la misma paradoja cuando esta se formula con el indexical, pues un cambio de agente garantiza un cambio en el contenido, por más que el carácter sea el mismo. Tómense los siguientes casos:

1. $Car[\phi] = (< t, w, p, a_1 > \rightarrow Con[\phi])$
2. $Car[\phi] = (< t, w, p, a_2 > \rightarrow Con[\phi])$
3. $Car[\phi] = (< t, w, p, a_3 > \rightarrow Con[\phi])$
4. $Car[\phi] = (< t, w, p, a_4 > \rightarrow Con[\phi])$

Donde ϕ abrevia 'estoy mintiendo', a_1 soy yo, a_2 es Graham Priest, a_3 es Jaakko Hintikka y a_4 es Britney Spears. En todos estos casos hay una instancia distinta de la paradoja debido al cambio de contexto. Y esto se cumple de manera general debido a las normas que rigen el carácter de oraciones autorreferenciales gracias al 'yo': para cualquier agente n, el contenido de ϕ se debe determinar en las circunstancias $< t, w, p, a_n >$, de tal modo que $Car[\phi] = (< t, w, p, a_n > \rightarrow Con[\phi])$

Pero con la diagonalización no sucede lo mismo. La paradoja es la misma si la escribe Priest, la escribe Hintikka o si la escribe Britney, etc. Esto es debido a que las oraciones del sistema **Q** no tienen un contenido variable, dado que en ellas no hay expresiones deícticas: el conjunto de enunciados derivables de **Q** no contiene ni indéxicos ni demostrativos debido a que estos no se pueden definir dentro del sistema. La diagonalización, contrario al indéxico, es un mecanismo *compartido* para lograr autorreferencia, varios agentes pueden expresar la misma paradoja acudiendo a la diagonalización por medio de cuantificadores.

Pero eso no es todo. Otro aspecto importante que se señaló en el capítulo 4 acerca del indéxico 'yo' es lo restringido que es su uso para referirme a cosas del mundo: usando 'yo' de manera deíctica no puedo hacer más que referirme a mí mismo (Kaplan, 1989b, p. 505). Al punto que, como se mencionó, cuando la paradoja se formula usando este indéxico el agente que emite la paradoja está en una posición privilegiada frente a la paradoja, dado que nadie más puede formular la misma paradoja. Como se explicó, esto significa que nadie puede instanciar la misma versión de la paradoja de otro agente.

Pero esto no sucede con enunciados como σ, en donde, si se fija la numeración de los elementos del sistema de manera distinta, con el enunciado σ se estaría haciendo referencia a cualquier otro enunciado cuando se diagonaliza bajo la nueva numeración. Es decir, el sistema es lo suficientemente flexible para permitir referenciar otros objetos del sistema. Por ejemplo, bajo una interpretación, un número de Gödel $\mathbf{n}$ puede hacer referencia a σ, pero bajo otra puede hacer referencia a otro enunciado del sistema distinto.

Así, con σ no podemos expresar lo mismo que se expresa con 'estoy mintiendo' en ningún caso. Esto se debe a que (1) dentro del sistema $\mathbf{Q}$ no podemos hacer referencia más que a enunciados propios del sistema y (2) la diagonalización mediante cuantificadores es un mecanismo compartido de autorreferencia (esto es, la referencia de un enunciado diagonalizado no cambia cada vez que en el contexto cambia el agente) con la cual se puede hacer referencia a más de un objeto; mientras que el indéxico 'yo' no es un mecanismo de autorreferencia compartido dado que el significado de una emisión con este indéxico varía cada vez que el agente cambia y tiene ciertos constreñimientos a la hora de referir, dado que el carácter de este indéxico solo nos permite referirnos a nosotros mismos.

5.2.2 ¿Podemos expresar 'esta oración es falsa' mediante σ?

Esta subsección es similar a la inmediatamente anterior. En aras de comparar y contrastar, tomo la referencia de ambas paradojas, las comparo y luego hago lo mismo con la forma en la que se logra autorreferencia. Las diferencias clave en este caso es que la versión de la paradoja es sobre oraciones y el modo de lograr autorreferencia es mediante el demostrativo 'esta'; mientras que, de nuevo, en lenguaje formal la paradoja es sobre un número que representa una oración y utiliza la diagonalización mediante cuantificadores para lograr autorreferencia. Para mayor claridad, recuérdese que las formulaciones son:

1. Esta [δ] oración es falsa;

2. $\exists x(x = \ulcorner a \urcorner \wedge \exists y(Diag(x, y) \wedge \neg \Gamma(y)))$

Surgen las mismas dos preguntas ¿puede hacerse referencia a lo mismo a lo que refiere (1) en un sistema como **Q**? y ¿puede la diagonalización como mecanismo autorreferencial cumplir el mismo papel que cumple 'esta' en una oración autorreferencial? Teniendo en cuenta lo expuesto en los capítulos 3 y 4 de este trabajo, creo que la respuesta a lo primero es sí, siempre y cuando se hagan algunas concesiones sobre las diferencias entre los conceptos de fórmula y oración; la respuesta a los segundo es que no, debido a las grandes diferencias entre el demostrativo y la diagnonalización. Antes de proceder, recuérdese que por motivos explicativos se toman las oraciones como las portadoras de la verdad.

Un aspecto que se señaló en el capítulo 3 es la naturaleza e importancia de los números de Gödel, en donde se remarcó que estos son herramientas para nombrar. Los números de Gödel están para nombrar expresiones del sistema. De manera que un número de Gödel $\ulcorner n \urcorner$ es el nombre de una expresión n del sistema.

El sentido de 'expresión' aquí es sumamente importante. Una consecuencia de que el sistema pueda representar únicamente las funciones recursivas es que dentro del sistema no se pueden representar fórmulas mal formadas mediante numeración de Gödel, ya que estas no son derivables dentro del sistema **Q**. Por lo tanto aquí 'expresión' significa fórmula bien formada, una secuencia de símbolos que cumple con las reglas de formación de un lenguaje formal particular (Cook, 2009, p. 312).

De tal manera que en σ lo que se está diciendo es que hay una fórmula bien formada designada mediante el nombre (número de Gödel) $\ulcorner \sigma \urcorner$ es falsa (mediante referencia con cuantificadores). Aquí es donde parece que hay que hacer una concesión.

Podemos aceptar que (1) y (2) hacen referencia a la misma clase de objetos solo si concedemos que funcionalmente las fórmulas y las oraciones son iguales. Es decir, una fórmula y una oración cumplen la misma función (ser portadores de la verdad), solo que una lo hace en lenguajes formales y la otra lo hace en lenguajes naturales.

Por ahora no estoy dispuesto a comprometerme con el hecho que las fórmulas y las oraciones sean funcionalmente equivalentes, pero tampoco tengo razones para rechazar esto. Tomar posición sobre esta materia supera los límites de esta investigación. Por mor del argumento voy a suponer que es cierto. Si no lo es, tanto peor para Priest.

Sin embargo, creo que podemos tener mucha mayor certeza respecto la otra pregunta, que es sobre la forma de lograr autorreferencia. Una de las dificultades de la semántica que se expuso en el capítulo 3 fue cómo dar cuenta de la referencia demostrativa y el rol que tenían tanto el contexto como las intenciones directrices.

Así, según la semántica de Kaplan-Perry una expresión como 'esta oración es falsa' debe analizarse en términos de su carácter estable, contenido variable e intención directriz. De esta manera, 'esta oración es falsa' es en realidad 'esta[δ] oración es falsa', para la cual la intención directriz δ hace parte del carácter y a la vez es necesario saber el contexto de emisión y circunstancias de evaluación para asignarle un contenido.

Con esto en mente, puedo hacer tantas instancias de la paradoja como se me antoje:

1. Esta[δ_1] oración es falsa.

2. Esta[δ_2] oración es falsa.

3. Esta[δ_3] oración es falsa.

n. Esta[δ_n] oración es falsa.

En estos casos, el subíndice de cada 'δ' representa el tipo de intención directriz. De modo que δ_1 representa la intención directriz según la cual un agente refiere a la oración 1 de la lista. Esto hace posible que, por ejemplo, el contenido se mantenga a pesar de que varíe el agente de emisión: dos personas pueden referirse a la misma oración por medio del mismo tipo de intención directriz si, por ejemplo, la primera afirma 'esta[δ_1] oración es falsa' y la segunda dice 'esa [δ_1] oración es falsa'.

De este modo, siempre que cambie el tipo de intención directriz, será una instancia distinta de la paradoja debido al cambio de carácter que implica el cambio de intención directriz (Kaplan, 1989a, pp. 587-589)). Dadas las reglas que rigen las intenciones directrices, para cualquier intención directriz δ_n que haga parte del carácter de 'esta oración es falsa', esta genera una nueva instancia de la paradoja siempre que δ_n difiera de cualquier intención directriz δ_m ya asociada con alguna emisión de 'esta oración es falsa'.

Lograr este tipo de reiteración es imposible con la diagonalización mediante cuantificadores, pues no importa cuántas veces escriba $\exists x(x = \ulcorner a \urcorner \wedge \exists y(Diag(x, y) \wedge \neg\Gamma(y)))$ o cuantas veces diagonalice σ cambiando de contexto o de circunstancia, esta fórmula nunca depende del contexto en el que sea emitida, por lo que es imposible generar instancias en donde la fórmula cambie de referencia, como sucedía con el cambio de intenciones directrices. Esto se debe precisamente a que la paradoja en lenguaje formal como tal no es una expresión dependiente del contexto, tiene tanto contenido como carácter estable.

Así, si no se hace una concesión respecto a la equivalencia entre fórmulas del lenguaje formal y oraciones del lenguaje natural, las paradojas son diferentes; si se hace, son también diferentes. A pesar de que podemos encontrar un punto común entre las dos formulaciones de la paradoja, no podemos expresar 'esta oración es falsa' por medio de σ en ningún caso. Esto se debe a que el poder expresivo del mecanismo de autorreferencia en un caso tiene una capacidad reiterativa que el otro mecanismo no tiene. Por lo tanto no es posible crear distintas instancias de la paradoja en lenguajes formales. Y de esta manera no tengo la misma paradoja si escribo $\exists x(x = \ulcorner \sigma \urcorner \wedge \exists y(Diag(x, y) \wedge \neg\Gamma(y)))$ que si escribo 'esta[δ] oración es falsa' con una intención directriz dirigida a esa misma oración.

5.3 Paradojas del mentiroso y el *Inclosure Schema*

Volvamos a Priest, quien en *Beyond the Limits of Thought* asegura que (1) "all the standard paradoxes of self-reference fit the [Inclosure] Schema" (Priest, 1995/2002, p. 277) y (2) que a pesar de que existen muchas variedades de la paradoja del mentiroso "[t]he exact *mechanisms* of reference are, in a clear sense, accidental to what is going on" (Priest, 1995/2002, p. 290), por lo que para él son la misma paradoja: "In all versions of the liar paradox, there is a sentence, σ, such that: σ is false iff $<\sigma>$ is false (1995/2002, p. 290). De estas dos afirmaciones debería seguirse que todas las variedades de la paradoja del mentiroso encajan (*fit*) con el *Inclosure Schema*. Sin embargo, lo dicho en este trabajo nos da razones para dudar esto último.

Según se comentó en los capítulos 2 y 3, cuando Priest explica cómo la paradoja del mentiroso cumple en el *Inclosure Schema* solo explica cómo una versión de la paradoja cumple con el esquema: la versión en lenguaje formal. Leyendo esto según la segunda tesis de Priest que se mencionó en el párrafo anterior es claro que para él no existe una diferencia real entre esta versión de la paradoja y las otras dos en lenguaje natural que él identifica, y que por lo tanto basta probar que la versión del lenguaje formal cumple con el esquema para pensar que las otras versiones también lo hacen.

Pero esto último parece ser falso. Este trabajo intentó dar buenas razones para pensar que si bien estas tres paradojas comparten algunos aspectos en cuanto su estructura, en realidad son paradojas suficientemente distintas, en las cuales los mecanismos para lograr autorreferencia y los objetos a los que hace referencia no solo son distintos, sino que son inexpresables entre sí. En su clasificación original de las paradojas, Priest comienza a distinguir paradojas que tienen la misma estructura precisamente por estos criterios (Priest, 1995/2002, p. 131), recuérdese el Cuadro 2.1 en el Capítulo 2, en el cual Priest distinguía paradojas con la misma estructura según el mecanismo de autorreferencia y al tipo de objetos a los que referían. Sin embargo, ahora a pesar de que tenemos tres paradojas que son distintas en estos aspectos, Priest insiste en tratarlas como la misma.

Esto tiene dos problemas. El primero de ellos es que Priest todavía no ha demostrado que “all the standard paradoxes of self-reference fit the [Inclosure] Schema” (Priest, 1995/2002, p. 277), hay dos paradojas las cuales no sabemos si cumplen con el esquema: ‘estoy mintiendo’ y ‘esta oración es falsa’.

A lo mejor sí lo cumplen, pues parece que tienen la misma estructura que otras paradojas, no solo para del mentiroso en lenguaje formal. Pero haría falta, por ejemplo, ver bajo qué totalidad (Ω) se deben evaluar las paradojas para que cumplan con el esquema.[56]

[56] Recuérdese que según el esquema debe existir una totalidad Ω tal que cuando se le aplica una función o diagonalizador $\delta(x)$ el resultado termina siendo un enunciado paradójico: el conjunto de todos los conjuntos que no pertenecen a sí mismos pertenece y no pertence a sí mismo (paradoja de Russell-Zermelo), es posible y no es posible pensar en el conjunto de todos los pensamientos (quinta

El otro problema es que el esquema podría llegar a trivializarse si se acepta la postura de Priest acerca de las paradojas del mentiroso. Inicialmente, Priest diferencia paradojas que tienen la misma estructura unas de otras por el mecanismo para lograr autorreferencia y los objetos sobre los cuales se hace referencia en la paradoja. Por ejemplo, vale la pena distinguir la paradoja de Russell-Zermelo de la paradoja de König porque en la primera hay autorreferencia en tanto se admite la existencia de conjuntos que no pertenecen a sí mismos y es sobre conjuntos; mientas que en la paradoja de König la autorreferencia se logra al intentar definir al primer ordinal no definible y la paradoja versa sobre números ordinales.

Con las paradojas del mentiroso tenemos algo similar. Paradojas que difieren en la forma en la que logran autorreferencia y que refieren a tipos de objetos distintos. Pero si nos tomamos enserio la posición de Priest, bajo una sola paradoja (σ) podríamos tener tres. Pero si esto es así ¿por qué detenerse allí? Si con σ capturo otras dos paradojas que son completamente distintas a σ misma, entonces con σ podría capturar el resto de paradojas de la autorreferencia. En este sentido todas las paradojas de la autorreferencia se reducirían a σ, en tanto no importaría ni el diagonalizador ni el tipo de objetos respecto de los cuales es posible formular la paradoja.

Pero esto es problemático, pues, para Priest, la paradoja que representa σ es distinta a la paradoja 'el conjunto de todos los conjuntos que pertenecen a sí mismos pertenece y no pertenece a sí mismo'. Por lo que se supone que aunque tienen la misma estructura (y por ende deberían tener la misma solución) debemos diferenciar entre estas dos paradojas. Pero Priest no le ofrece este mismo tratamiento a las tres paradojas del mentiroso, lo cual puede llevar al problema que se acaba de exponer.

Si no necesitamos distinguir entre las paradojas, no valdría la pena ni siquiera decir que hay una solución uniforme, sino que todas las paradojas se resuelven de exactamente la misma manera, porque serían la misma paradoja. Por ejemplo, Priest mismo dice que una manera aislada de resolver su versión de la paradoja del mentiroso en lenguaje formal es apelando al nominalismo matemático (Priest, 1995/2002). Dado que esta es la

antinomia), hay un número r que es y no es definible (paradoja de Richard), etc. En el primer caso Ω es el conjunto de todos los conjuntos que pertencen a sí mismos, en el segundo es el conjunto de todos los pensamientos y en el tercero el conjunto de números reales definibles.

solución a una paradoja que supuestamente captura todas las paradojas de la autoreferencia (σ), entonces el nominalismo matemático es la solución a todas las paradojas de la autorreferencia y no hay necesidad de formular, desarrollar o defender el dialeteísmo.

De esta manera, según Priest, todas las versiones de la paradoja del mentiroso encajan en el *Inclosure Schema* en tanto σ lo hace. Lo cual es falso, no se trata de versiones distintas de la paradoja, sino de paradojas distintas. Pero de ahí no se sigue debemos rechazar el dialeteísmo. Deberíamos rechazarlo solo en el caso que la objeción planteada aquí fuera insuperable. Pero no lo es.

El dialeteísmo de Priest es una posición filosófica que podría aceptar las conclusiones de este escrito sin perder mucho, solo hace falta hacer una serie de concesiones. La primera de estas es aceptar que los mecanismos con los que se logra referencia (*the exact mechanisms of reference*) no son accidentales, pues, por ejemplo, el uso del indéxico y el uso de demostrativo en enunciados autorreferenciales paradójicos no son expresables mediante la diagonalización.

La segunda concesión es que precisamente no hay tal cosa como *la* paradoja del mentiroso, sino varias paradojas del mentiroso (una con indéxico, una con demostrativo, una con diagonalización, por decir algunas), las cuales ameritan un análisis más detallado. Este análisis sería precisamente una demostración de cómo las dos paradojas del mentiroso en lenguaje natural encajan en el *Inclosure Schema*.

En tercer lugar, la última concesión que tendría que hacer el dialeteísmo de Priest es prestar más atención a la forma en la que se formulan las paradojas del mentiroso, de modo que no se suponga que una se reduce a otra simplemente porque tratan con la noción de verdad. Dentro de la literatura es fácil encontrar otras paradojas que son similares a las paradojas en el lenguaje natural tratadas en este trabajo, pero que en últimas son claramente diferentes.

La primera paradoja es la denominada la paradoja de la postal, en la cual se encuentra una postal que dice de un lado 'la oración del otro lado es verdadera' y al otro lado dice 'la oración del otro lado es falsa' (O'Connor & Robertson, 2005). Esto crea una auténtica paradoja en tanto evaluar el valor de verdad de cualquiera de las dos caras de la postal acaba en contradicción. Aquí vemos una paradoja en la que las oraciones hacen referencia a oraciones pero la forma de lograr autorreferencia es mediante descripciones (el x tal que Fx) e indirecto, la oración no refiere directamente a sí misma, sino a otra oración que refiere directamente a la primera. Solo con esto ya valdría la pena preguntarse si es posible, mediante alguna de las paradojas del mentiroso estudiadas en este trabajo, reducir la paradoja de la postal a una de estas paradojas.

Algo similar sucede con versiones de la paradoja del mentiroso que logran contradicciones en su valor de verdad refiriendo mediante nombres: considérese una oración llamada *Paradoja*, la cual dice que es falsa: Paradoja $=_{def.}$ Paradoja es falsa. Solo el hecho que se refiera mediante nombres debería por lo menos dudar que esta paradoja no puede reducirse a otra que utilice otro mecanismo de autorreferencia. Y lo mismo sucede con otras paradojas del mentiroso en lenguaje formal como la de Kripke (ver nota 16). Así, esta última concesión consiste en tener más cuidado cuando se esté tratando una paradoja del mentiroso, no se debe suponer que se puede reducir a otra versión de la misma paradoja, sino que se debe preguntar si dado su mecanismo para lograr autorreferencia y el tipo de objetos a los que refiere esta paradoja es expresable en términos de otra.

Capítulo 6

Conclusión

"If one uses Gödel's Theorem as a metaphor, as a source of inspiration, rather than trying to translate it literally into the language of psychology or of any other discipline, then perhaps it can suggest new truths in psychology or other areas. But it is quite unjustifiable to translate it directly into a statement of another discipline and take that as equally valid. It would be a large mistake to think that what has been worked out with the utmost delicacy in mathematical logic should hold without modification in a completely different area."

(Hofstadter, 1979/1999, p. 696)

Este trabajo termina presentando algunas conclusiones, lo que en esta ocasión significa que se exploran algunas consecuencias de lo defendido a lo largo de la tesis, se abordan limitaciones de la misma y se mencionan algunos puntos en los que vale la pena realizar investigaciones en el futuro.

6.1 Consecuencias

La primera consecuencia que se sigue de inmediato del análisis hecho en los capítulos anteriores es el hecho que 'estoy mintiendo' y 'esta oración es falsa' son paradojas distintas. Esto se debe, en primer lugar, al tipo de objetos al que refiere una y otra paradoja: una refiere a expresiones dichas en un contexto particular por un agente particular mientras que la otra paradoja refiere a oraciones. Sin embargo, la diferencia más relevante ocurre cuando se comparan los mecanismos de autorreferencia. Como se mencionó en el capítulo 4, es posible defender que no se puede expresar con ninguna otra expresión aquello que está expresado mediante el deíctico ‘yo’.

Como se comentó en su momento, la forma particular en la que esto es relevante para las expresiones en lenguaje natural es respecto a la imposibilidad de reemplazar el ‘yo’ con ‘esta’. Esto se debe, según Perry, a que a pesar de que quiera hablar de mí utilizando una expresión como ‘esta persona’, puede que no sepa que me estoy refiriendo a mí, como sucedía en el ejemplo del azúcar que se regaba en el supermercado; en cambio, cuando utilizo ‘yo’ para referirme a mí mismo no hay duda de que me refiero a mí mismo (Perry, 1993a). De modo que si antes teníamos razones para afirmar que la paradoja en lenguaje formal es diferente a las paradojas en lenguaje natural, ahora también podemos afirmar que las dos paradojas en lenguaje natural son distintas entre sí.

De manera similar, podría darse una multiplicación de paradojas del mentiroso en caso de que se acepten las de las diferencias aquí planteadas. Por ejemplo, podríamos combinar objetos de referencia y mecanismos de autorreferencia en lenguaje natural para obtener enunciados que de manera preliminar pueden constituir otras paradojas del mentiroso. Por ejemplo, habría diferencias importantes entre una paradoja como ‘yo estoy mintiendo’ y ‘la oración que estoy diciendo es falsa’ dado el tipo de objetos al que refieren, a pesar de que la forma de lograr autorreferencia (el deíctico ‘yo’).

La segunda conclusión tiene que ver con el lugar de las paradojas del mentiroso en el dialeteísmo. Si bien las concesiones que el dialeteísmo de Priest debe hacer para dar cuenta de las paradojas del mentiroso son pocas, esta nueva versión del dialeteísmo tendría que justificar que las otras paradojas del mentiroso encajan en el *Inclosure Schema*. Esto se debe a que rechazar que estas paradojas sean la misma nos obliga a interpretar la prueba de que la paradoja del mentiroso encaja en el esquema como una prueba solo para la paradoja en lenguaje formal.

Una manera de hacer esto puede ser tomando la versión estándar del *Inclosure Schema* presentada en el segundo capítulo. Sin embargo esto puede resultar poco provechoso para las paradojas en lenguaje natural debido a que estas no están expresadas en el lenguaje de la teoría de conjuntos en las que están formulados tanto el *Inclosure Schema* como la paradoja en lenguaje formal. En cambio, se podría expresar una versión equivalente el esquema en términos de propiedades: Dada una propiedad ϕ y una propiedad de propiedades Ψ se tiene que si

1. ϕ existe y $\Psi(\phi)$ (Existencia)

2. Y θ tal que $\Psi(\theta)$ y $\forall x(\theta x \rightarrow \phi x)$, entonces:

   (a) $\neg\theta\delta(\theta)$ (Transcendencia)

   (b) $\phi\delta(\theta)$ (Clausura) (Priest, 1995/2002, p. 280 (pie de página 26)).

De esta manera, una primera aproximación a cómo las paradojas del mentiroso en el lenguaje natural encajan en el esquema podría darse reemplazando los valores del esquema por los valores relevantes en cada caso. Esto, por supuesto, exige ampliar el lenguaje formal utilizando mucho más allá del sistema **Q** utilizado en el capítulo 3 para poder dar cuenta del comportamiento de expresiones dependientes del contexto. Lamentablemente, esta tarea supera ampliamente los límites de esta investigación.

6.2 Limitaciones

La limitación más evidente de este trabajo es que, como se dijo en su momento, en *Beyond the Limits of Thought*, Priest menciona hasta cinco formulaciones distintas de la paradoja del mentiroso, unas en lenguaje formal, otras en lenguaje natural. Sin embargo, para analizarlas se hubiera necesitado mucho más espacio. Esto significa que a pesar de que reconozco que quedan más versiones de la paradoja del mentiroso por estudiar, espero que los instrumentos expuestos aquí sirvan como insumo para determinar cuáles de estas son versiones de una paradoja o auténticas paradojas.

Otra limitación de la tesis es la cantidad de nociones que da por supuestas. Esto es especialmente cierto para los conceptos de filosofía del lenguaje utilizados. Conceptos como referencia, significado, autorreferencia, actitud, oración, expresión, proposición, entre otras, son algunos conceptos de gran importancia a lo largo del trabajo de los cuales no se hizo una discusión a profundidad.

Esto me obligó en muchas oportunidades a tener que hacer varias concesiones por el bien del trabajo, como aceptar que las oraciones son las portadoras de la verdad, que en ciertos contextos podría llegar a pensarse que las fórmulas y las oraciones son equivalentes en cuanto a su función o que mentir en el contexto de la paradoja significa simplemente decir algo falso, sin entrar a discutir el concepto de la mentira o el mentir como una acción. Estas concesiones se debieron hacer dado que muchas veces las discusiones en torno a los conceptos en cuestión superaban ampliamente el ámbito de la investigación.

Unido a esto viene el hecho de que cuando sí se tomó partido por alguna posición sobre aspectos filosóficos, las elecciones fueron sumamente convenientes. La teoría de Kaplan pretende ver el lenguaje de manera formal dada a la influencia que Carnap y Montague ejercen sobre él y en este sentido la teoría de Picollo se ajusta muy bien a esta manera de ver el lenguaje. Dado esto, una clara objeción a este trabajo es el hecho que lo que se ha argumentado aquí solo funciona dada una concepción muy particular de los deícticos y la referencia en lenguaje formal. Y aún cuando parece que estas teorías con compatibles entre sí, estas no fueron diseñadas para soportar una análisis dialeteísta de las paradojas de la autorreferencia, algo que este trabajo supone y no entra a discutir, pero puede ser fuente de críticas al nivel de fundamentación.

De esta manera (1) si alguno de estos dos aspectos cambia, es muy probable que se lleguen a conclusiones muy distintas y (2) los resultados de este trabajo dependen de teorías filosóficas que siguen abiertas a debate, son relativamente nuevas en los círculos filosóficos y con las cuales muchas personas se encuentran en profundo desacuerdo. Por lo que si se rechaza o bien la teoría de Picollo o bien la de Kaplan, este trabajo no tendría ningún sentido, pues requiere de ambas para tenerlo.

Esto último es particularmente cierto del dialeteísmo. Si bien esta no es una tesis acerca de la verdad del dialeteísmo, por lo menos se supone la plausibilidad del mismo, y simplemente señala un problema que puede ser solucionado. De este modo, alguien que considere que el dialeteísmo es falso también está rechazando no solo que lo afirmado en este trabajo sea verdadero, sino incluso que tenga sentido, pues está corrigiendo un error de un teoría filosófica que es falsa de plano.

Esto hace que el público al que va dirigido este trabajo sea sumamente reducido, pues quienes rechazan el dialeteísmo no tendrían en principio interés en leerlo y quienes aceptan el dialeteísmo puede que rechacen este trabajo o bien porque tienen otra teoría del significado o bien porque consideran que los desafíos planteados triviales. Así, este trabajo en últimas está dirigido únicamente o bien a dialeteístas que consideren que el dialeteísmo puede ser mejorable al nivel de fundamentación o bien a personas para las cuales el dialeteísmo es una teoría por lo menos plausible y quieran entender mejor las motivaciones para este.

6.3 Posibles investigaciones futuras

Como se adelantaba hace unos pocos párrafos, el lugar de las paradojas del mentiroso en el dialeteísmo es un asunto que amerita mayor investigación al respecto. Especialmente porque pienso que una vez que el dialeteísmo pueda explicar cómo las paradojas del mentiroso son paradojas de la clausura (*inclosure paradoxes*, como diría Priest) es posible robustecer incluso más la motivación del dialeteísmo explorada aquí, según la cual el dialeteísmo es la única posición que ofrece una solución uniforme a las paradojas de la autorreferencia al tratarlas como contradicciones verdaderas.

Si lo dicho hasta este momento es cierto, entonces es posible que muchas de las soluciones a las paradojas del mentiroso tengan un alcance bastante reducido. Esto, pues muchas veces estas soluciones se han planteado usando una u otra paradoja del mentiroso, sin tener en cuenta que se trata de auténticas paradojas del mentiroso, lo que genera que las soluciones estén planteadas para una única paradoja. Por el contrario, si se toma la solución uniforme que ofrece el dialeteísmo no es necesario molestarse con una lista cada vez más larga de paradojas del mentiroso y soluciones que las acompañan.

El siguiente punto tiene que ver con la referencia y el significado en los lenguajes formales. El primero de estos puntos es qué es lo que realmente dicen las oraciones de Gödel y, de manera más general, qué es lo que dicen las oraciones que utilizan la diagonalización mediante cuantificadores o métodos similares de autorreferencia. Aquí con 'lo que dicen' me refiero, si se quiere, al contenido de estas expresiones. Como se explicó en el capítulo 3, la concepción heredada es que son expresiones deícticas, lo cual se demostró que no es el caso. En un lado más moderado de la discusión, sin embargo, se encuentran aquellos que afirman que se tratan de descripciones definidas. Como se indicó en su momento, esto difícilmente puede considerarse cierto. No me queda otra opción que dejar la pregunta abierta y esperar poder responderla en el futuro.

Pienso que la importancia de este último asunto no debe subestimarse. Si en este momento no sabemos qué es lo que dice exactamente una oración que logra autorreferencia mediante diagonalización, eso significa que, por ejemplo, todavía no entendemos plenamente el primer teorema sobre la incompletitud de la aritmética, v.gr. no sabemos qué es lo que dice una oración que dice de sí misma que es indecidible. Un estudio más juicioso de la autorreferencia, y en general de la referencia, en lenguajes formales no solo nos puede llevar a entender mejor este tipo de oraciones, sino también entender de manera adecuada oraciones como las oraciones de Henkin (oraciones que dicen de sí mismas que son decidibles) y en general la relación que existe entre los lenguajes formales y el lenguaje natural.

Apéndice A

Paradojas de la autorreferencia en *Beyond the Limits of Thought*

Uno de los ejemplos más sencillos de las paradojas es la siguiente: Pinocho es una marioneta de madera, si Pinocho dice algo que no es verdad, su nariz crece. Supongamos que Pinocho dice 'Mi nariz está creciendo' ¿debe la nariz de Pinocho crecer? Supongamos que la nariz de Pinocho crece, si crece, es porque dijo algo que no era verdad, pero si crece, lo que dijo es verdad, pues en efecto su nariz está creciendo, dado que lo que dijo es verdad, su nariz no debería crecer, pero si no crece es porque dijo algo verdadero, lo cual es falso, porque si su nariz no crece, lo que dijo es falso, entonces la nariz crece y no crece.

Pero si suponemos que no crece, es porque Pinocho dijo la verdad, pero esto es falso, porque su nariz no está creciendo, como dijo algo falso, su nariz debe crecer, pero si su nariz crece, entonces habría dicho la verdad y su nariz no debería crecer, entonces su nariz crece y no crece.[57] En ambos casos la nariz de Pinocho debe crecer y no debe crecer. Por lo que se concluye precisamente que la nariz de Pinocho crece y no crece, esta es una conclusión que, aunque contradictoria, se sigue de la verdad de las premisas pero las contradice. Por ende, se trata de una paradoja.

Esta es una de muchas paradojas. En esta sección expongo con detalle las llamadas

---

[57] Esta es una versión de la Paradoja de Pinocho, formulada por Peter Elridge-Smith y Veronique Eldridge-Smith, para más información ver Eldridge-Smith & Eldridge-Smith (2010).

paradojas de la autorreferencia a las que se hizo referencia en el capítulo 2. La autorreferencia es un fenómeno según el cual una cosa se refiere a sí misma o a una parte de sí misma, puede ser una fotografía, un dibujo o una oración (Bolander, 2017). Lograr autorreferencia en oraciones es sencillo, pero hay ciertas oraciones auto-referenciales que llevan a paradojas, exponer un número significativo de estas paradojas es el objetivo de este apéndice

A.1 Las paradojas de la autorreferencia

Este apartado tiene como objetivo dejar claras las paradojas de la autorreferencia expuestas en un lenguaje lo menos simbólico posible. Esta sección sirve como referencia para el primer capítulo, en donde se utiliza un aparato lógico que puede resultar complejo.

Si bien en esta sección se exponen y analizan varias paradojas de la autorreferencia, no sobra decir que la lista no es exhaustiva.[58]

A.1.1 La paradoja de Russell-Zermelo

Esta paradoja tiene que ver con la noción de pertenencia. Fue formulada independientemente por Bertrand Russell y Ernst Zermelo. Russell la desarrolló antes de junio de 1902 y se la hizo saber a Gottlob Frege en una carta dirigida a este del 16 de junio de 1902 (Russell, 2016, pp. 575-576), la paradoja fue publicada en 1903 en *Principle of Mathematics* (Russell, 1903, §344). Por su parte, Zermelo desarrolló independientemente el mismo resultado y compartió sus resultados con Edmund Husserl y David Hilbert hacia 1902 (Rang & Thomas, 1981).

La paradoja es la siguiente. Supongamos que queremos agrupar grupos de cosas, por ejemplo, podemos agrupar a todos los filósofos antiguos en un mismo grupo, lo mismo con los números primos, los libros, etc. Estos grupos están constituidos por elementos: el grupo de los filósofos antiguos tiene a Pirrón, Hipatía, Carneades, Diógenes, etc.; el de los números primos contiene a 2, 3, 5, 7 ...; y así sucesivamente. A estos 'grupos' se

[58] Para una un análisis más detallado de estas, otras paradojas de la autorreferencia y su discusión durante el siglo XX, ver Cantini (2009).

les denomina conjuntos. Estos conjuntos a su vez pueden agruparse en grupos, creando conjuntos de conjuntos. Por ejemplo, al conjunto de todos los filósofos pertenece el conjunto de todos los filósofos antiguos, pero también el conjunto de los filósofos medievales, el conjunto de los filósofos contemporáneos y muchos otros.

Tomando la noción de conjunto de conjunto podemos entender nociones más complicadas. Hay conjuntos que no pertenecen a sí mismos. Por ejemplo, el conjunto de todos los verbos gramaticales no es un verbo gramatical y por lo tanto no pertenece a sí mismo. Sin embargo, suponiendo que exista un conjunto de todos los conjuntos que no pertenecen a sí mismos, ¿este conjunto pertenece o no pertenece a sí mismo? Aquí es donde la paradoja aparece (recordemos que debe tratarse de un argumento válido que lleve a una contradicción y además debe existir autorreferencia).

La pregunta tiene dos posibles respuestas, o bien pertenece a sí mismo o no pertenece a sí mismo. Si pertenece a sí mismo, entonces pertenece al conjunto de conjuntos que no pertenecen a sí mismos, entonces, en efecto, no pertenece a sí mismo, porque precisamente hace parte de los conjuntos que no pertenecen a sí mismos, por lo tanto pertenece a sí mismo y no pertenece a sí mismo. Pero si se supone que no pertenece a sí mismo, entonces no pertenece al conjunto de todos los conjuntos que no pertenecen a sí mismos, por lo cual es un conjunto que pertenece a sí mismo. Es, por lo tanto, un conjunto que pertenece a sí mismo y no pertenece a sí mismo.

En cualquiera de los dos casos se obtiene la misma contradicción, el conjunto pertenece a sí mismo y no pertenece a sí mismo. Por lo tanto, la conclusión es que el conjunto de todos los conjuntos que no pertenecen a sí mismos pertenece a sí mismo y no pertenece a sí mismo.

A.1.2 La paradoja de Burali-Forti

Esta paradoja fue formulada por el matemático italiano Cesare Burali-Forti en 1897. El propósito de esta paradoja era mostrar que un sistema que permita construir el conjunto de todos los números ordinales lleva a contradicción (Burali-Forti, 1897, p. 154), lo cual prueba la existencia de números transfinitos, números que son mayores a cualquier número natural.

En matemáticas existe un grupo de números llamado los números naturales, son los números enteros positivos a partir del 0 (ISO, 2009). Estos números pueden utilizarse para diversidad de actividades, entre estas contar (uno, dos, tres...) y ordenar o enumerar (el primero, el segundo, el tercero...). Cuando se utilizan para contar, se denominan números cardinales, cuando se utilizan para enumerar, se les llama números ordinales.

La paradoja de Burali-Forti tiene que ver con el conjunto de números ordinales. Específicamente, con el conjunto de todos los números ordinales, al cual denominaremos On. Si x es un elemento de On y y es un elemento de x, entonces y es un elemento de On. Además, contamos con que todos los elementos de On están organizados de menor a mayor y hay un elemento que es el menor de todos.

Si lo anterior es correcto, entonces On es un ordinal, es decir pertenece a sí mismo. La razón es la siguiente, si los elementos de On están organizados de mayor a menor, entonces On es el mayor ordinal del conjunto de ordinales, pues contiene a todos los ordinales. Tomemos, por ejemplo x, el mayor ordinal dentro de On. Dado que On es un conjunto ordenado y x pertenece a On, entonces x es menor que On, pues On contiene a x. Dado esto, On pertenece al conjunto de ordinales, pues es mayor que cualquier otro ordinal

Si nos preguntamos si existe un sucesor de On (el menor ordinal mayor que On), encontramos que este es y no es un número ordinal. Es un número ordinal en tanto se utiliza para enumerar el conjunto de todos los número ordinales, es decir, es mayor que cualquier número ordinal y por ende contiene a todos los números ordinales; pero al mismo tiempo, no puede ser un número ordinal, dado que está por fuera del conjunto de todos los ordinales.

A.1.3 La paradoja de Mirimanoff

Explota los resultados de las paradojas de Russell-Zermelo y Burali-Forti para demostrar que la noción de totalidad es problemática, pero esta vez mediante una nueva herramienta (Mirimanoff, 1917, p. 52). En esta ocasión se trata de (1) el conjunto potencia y (2) el concepto de conjuntos bien formados y la totalidad de conjuntos bien formados, que en este caso denomino V.

Para poder explicar esta de forma adecuada se necesita entender la noción de conjunto bien fundado. Un conjunto está bien fundado cuando tiene un límite inferior, es decir, posee un primer elemento. Por otro lado, $\mathrm{P}(x)$ representa el conjunto potencia de x, es decir, el conjunto de todos los conjuntos que se pueden formar a partir de los elementos de x, $\mathrm{P}(x)$ incluye al conjunto vacío y a x mismo, pero $\mathrm{P}(x)$ no pertenece a x. Si x es un conjunto bien fundado, $\mathrm{P}(x)$ es un conjunto bien fundado.

Por ejemplo, suponiendo que x es el conjunto de números $\{1, 2\}$, el conjunto potencia de de x es $P(\mathrm{x}) = \{\{\emptyset\}, \{1\}, \{2\}, \{1, 2\}\}$. En este caso, tanto x como $\mathrm{P}(x)$ son conjuntos bien fundados, pues en ambos casos tienen un límite inferior, $\{1\}$ y $\{\emptyset\}$ respectivamente.

Volvamos a V, el conjunto todos los conjuntos bien fundados. Si tomamos $\mathrm{P}(V)$, la paradoja que surge es que $\mathrm{P}(V)$ no pertenece a V, pues es su conjunto potencia, pero a su vez sí pertenece a V, pues $\mathrm{P}(V)$ es un conjunto bien fundado.

A.1.4 La Quinta Antinomia de Kant

Continúo con la *Quinta Antinomia de Kant*. Para cualquier lector de Kant, hablar de la quinta antinomia puede resultar una imprecisión, en tanto en la *Crítica de la Razón Pura* (KrV) Kant sólo menciona cuatro antinomias y en el resto del *corpus* no hay referencia a una quinta antinomia. La *Quinta Antinomia* no fue propuesta por Kant, pero está inspirada en la línea de pensamiento de las cuatro antinomias, y por eso recibe este nombre. Fue formulada por Graham Priest en su libro *Beyond the Limits of Thought* (1995).

La *Crítica de la Razón Pura* está divida en la *Doctrina trascendental de los elementos* y en la *Doctrina trascendental del método*. En la *Doctrina trascendental de los elementos* se establecen cuáles son los elementos *a priori* de la sensibilidad y del entendimiento, Kant dedica la *Estética Trascendental* a los elementos *a priori de la sensibilidad*, mientras que la *Lógica Trascendental* analiza los elementos *a priori* del entendimiento. De esta manera, la *Lógica Trascendental* se divide nuevamente en la *Analítica trascendental* y en *Dialéctica Trascendental*; mientras que la *Analítica* se encarga de especificar la deducción y los usos apropiados de los conceptos puros del entendimiento, la *Dialéctica* demuestra los usos incorrectos de los conceptos expuestos en la *Analítica*, es decir, juicios a los que llegamos porque la razón va más allá de la experiencia posible, ya sea porque se llega a estos por inferencias formalmente incorrectas o porque se llega a dos argumentos formalmente válidos pero contradictorios entre sí acerca de la naturaleza del mundo. Si sucede lo primero, se trata de un paralogismo de la razón pura (A341/B399-A405/B432), si es lo segundo, es una antinomia de la razón pura o un conflicto de ideas trascendentales (KrV, A405/B433-A462/490).

Estas antinomias se presentan normalmente como tesis y antítesis. Por ejemplo, la primera antinomia o conflicto se expone de la siguiente manera en la KrV.

- *Tesis*: El mundo tiene un límite en el tiempo y en espacio: tiene un comienzo y está igualmente encerrado entre límites (KrV, A427/B455).
- *Antítesis*: El mundo no tiene un límite ni en el espacio ni en el tiempo: es infinito respecto del tiempo como del espacio (KrV, A427/B455).

La quinta antinomia es la siguiente (Priest, 1995/2002, pp. 100-101): por definición, el pensamiento que se tiene sobre un objeto es distinto al objeto mismo, si pienso en mi ejemplar de *Cien años de soledad*, mi pensamiento acerca de ese libro es distinto al libro mismo. Tomemos el conjunto de todos los pensamientos, ¿puedo pensar en este conjunto?

Si puedo pensar en este conjunto, entonces el pensamiento que tengo sobre el conjunto de todos los pensamientos es distinto al conjunto de todos los pensamientos, por definición. Pero si esto es así, entonces hay un pensamiento que no hace parte del conjunto de todos los pensamientos (el pensamiento sobre el conjunto de todos los pensamientos), de manera que no podría pensar en el conjunto de todos los pensamientos. Supongamos entonces que no puedo pensar en el conjunto de todos los pensamientos. Si esto es así, entonces acabo de pensar en el conjunto de todos los pensamientos, pues acabo de decir que no puedo pensarlo, para poder decir esto tuve que haber pensado en este, por lo que sí podría pensar en el conjunto de todos los pensamientos

De manera más kantiana:

- *Tesis*: Es posible pensar en el conjunto de todos los pensamientos.
- *Antítesis*: No es posible pensar en el conjunto de todos los pensamientos.

A.1.5 Las paradojas de König, Berry y Richard

Vale la pena agrupar estas paradojas en un mismo grupo por una razón: tienen que ver con la definibilidad de los números. La definibilidad es una noción importante tanto en matemáticas como en filosofía, algo es definible si y solo si existe una frase que se refiera a este y sólo a este, y que el que la frase refiera a este objeto no depende del contexto en el que sea usada, si quiero definir el número 2, puedo utilizar la oración "El único número natural que es par y es primo" (Priest, 1995/2002, p. 131). De esta manera, la paradoja de König trata sobre números ordinales indefinibles, la paradoja de Berry sobre números naturales definibles y la paradoja de Richard refiere a la definibilidad de los números reales.

La paradoja de König es la siguiente. El español es un idioma con un vocabulario finito, por lo tanto, en el español existe un número finito de definiciones que se crean combinando las palabras de su vocabulario. No obstante, los número ordinales son un conjunto infinito de números, por lo que es imposible definir todos y cada uno de estos. Dado un vocabulario finito y un conjunto infinito de números, entonces debe haber un primer número ordinal no definible, que, por definición, sería indefinible. Sin embargo, este número acaba de definirse como el primer número ordinal no definible (König, 1905).[59]

La paradoja de Berry también hace uso de la noción de definibilidad, en este caso tomemos el conjunto de número naturales definibles en menos de 19 palabras, tomemos el primer número natural que no se puede definir en menos de 19 palabras, este, por definición, no pertenece al conjunto de números naturales definibles en menos de 19 palabras, sin embargo este número acaba de ser definido en menos de 19 palabras, pues se definió como 'el primer número natural que no se puede definir en menos de 19 palabras', se necesitaron 14 palabras para hacerlo.[60]

---

59 Según Priest, hay una paradoja inspirada en el trabajo de Berkeley que es exactamente igual a la paradoja de König, salvo en que en la primera lo que genera la paradoja es un objeto concebible y no concebible: supongamos un que un objeto existe sin que sea concebido, existe por sí mismo; pero este objeto ya fue concebido, lo acabamos de hacer. Por lo tanto este objeto puede no ser concebido y puede ser concebido. Al respecto, ver Priest (1995/2002, secs. 4.4-4.9).

60 La paradoja de Berry apareció por primera vez en (Russell, 1906a, p. 345), Russell afirma que esta le fue sugerida por Mr. G.G. Berry, un bibliotecario de la Bodleian Library de la Universidad de Oxford. Esta versión en castellano no es exactamente igual a la original debido a la traducción que fue

La paradoja o antinomia de Richard también aprovecha la noción de definibilidad. Tómese todas las oraciones del español que definan un número real entre 0 y 1. Esta lista de oraciones debe organizarse alfabéticamente. La lista es infinita y se puede corresponder cada oración con un número real, de modo que la primera oración corresponde al primer número real, la segunda al segundo, etc. de modo r_n representa el número r que corresponde a la expresión número n . Ahora de definirá un número r de la siguiente manera: la parte entera de r es 0, el n-ésimo decimal de r es 1 si el n-ésimo decimal de r_n no es 1, y el n-ésimo decimal de r es 2 si el n-ésimo decimal de r_n es 1 (Richard, 1905, p. 545).

La expresión del párrafo anterior que va desde "la parte entera" hasta "r_n es 1" es una expresión del español que define r de manera no ambigua, por lo tanto, r debería ser uno de los números definidos en la lista, sin embargo, r se construyó de manera que no fuera igual a ningún r_n (o sea, de manera que no hubiera ninguna definición de la lista que le aplicara). Por lo tanto r es un número definible (pues se construyó una definición para ese número) y no es un número definible, pues no hace parte del conjunto de números definibles.

A.1.6 Las paradojas del mentiroso y del conocedor

Dado que anteriormente ya se expuso lo suficiente la paradoja del mentiroso, aquí vale la pena hacer no más que un breve comentario sobre esta y una paradoja muy similar (la pradoja del conocedor, *knower's paradox*). Una de las versiones de la paradoja del mentiroso es la siguiente: tómese la oración 'esta oración es falsa', dado que todas las oraciones deben ser verdaderas o falsas, el valor de verdad de esta oración debe poder ser determinado. Sin embargo, si la oración es verdadera, entonces es cierto que es falsa, por lo tanto la oración es falsa; pero si la oración es falsa, entonces es falso que la oración sea falsa, la oración debe ser verdadera. Por lo tanto, la oración es verdadera y falsa. Lo mismo sucede con la oración "esta oración no es conocida". Supongamos que conozco que esta oración es conocida, si es conocida entonces conozco que no conozco esa oración, por

necesario hacer algunos ajustes, pero el espíritu de la paradoja se mantiene. Vale la pena mencionar que es posible que Bebbo Levi haya descubierto la misma paradoja de manera independiente en 1908, ver (Levi, 1908) lo tanto no conozco la oración. Es decir, conozco y no conozco la oración.[61]

A.1.7 La paradoja de Grelling-Nelson

Por último, está la paradoja de Grelling-Nelson o del predicado heterológico. Esta paradoja fue descubierta en conjunto por Kurt Grelling y Leonard Nelson[62] en 1907, a veces incorrectamente atribuida a Hermann Weyl,[63] filósofo y matemático alemán que menciona una conocida paradoja derivadada de la paradoja de Russell (Weyl, 1918, p. 2).

Tomemos todos los predicados del español, los organizaremos en dos grupos, si el predicado le aplica a la palabra que designa el predicado, entonces lo pondremos en el grupo de predicados autológicos, de lo contrario, lo pondremos en el grupo de predicados heterológicos. Por ejemplo, 'pentasílabo' es un predicado autológico porque la palabra 'pentasílabo' es pentasílaba; 'rápido' es un predicado heterológico porque la palabra 'rápido' no es rápida. Se debe repetir este proceso con todos los predicados

Cuando llega el momento de clasificar el predicado 'heterológico' aparecen los problemas. Si 'heterológico' fuera autológico entonces la palabra 'heterológico' no sería heterológica, pero si es autológica, entonces le aplica el predicado que designa, por ende, la palabra 'heterológico' es heterológica; si 'heterológico' fuera heterológica entonces no sería autológica, pero si es heterológica, entonces no le aplica el predicado que designa, por ende, la palabra 'heterológico' es autológica, y he ahí la paradoja.

61 Esta última paradoja fue descubierta por David Kaplan y Richard Montague (1960).

62 Ver Grelling & Nelson (1907).

63 Ver, por ejemplo, Ramsey (1925, p. 20).

Referencias

Ballarin, R. (2017). Modern Origins of Modal Logic. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017). Metaphysics Research Lab, Stanford University.

Barwise, J. & Etchemendy, J. (1987). *The Liar: An Essay on Truth and Circularity*. Oxford University Press USA.

Beall, J., Glanzberg, M. & Ripley, D. (2019). Liar Paradox. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University.

Bergmann, M., Moor, J. & Nelson, J. (1980/2014). *The Logic Book* (6ª ed.). McGraw-Hill (Publicado originalmente en 1980).

Berto, F. (2009). *There's Something About Gödel: The Complete Guide to the Incompleteness Theorem*. Wiley-Blackwell.

Bobenrieth Miserda, A. (1996). *¿Inconsistencias? ¿Por qué no? Un estudio filosófico de la lógica paraconsistente*. Tercer Mundo Editores.

Bobenrieth Miserda, A. (2010). The Origins of the Use of the Argument of Trivialization in the Twentieth Century. *History and Philosophy of Logic*, *31*, 111-121.

Bolander, T. (2017). Self-Reference. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017). Metaphysics Research Lab, Stanford University.

Boolos, G. S., Burgess, J. P. & Jeffrey, R. C. (1974/2007). *Computability and Logic* (5ª ed.). Cambridge University Press (Publicado originalmente en 1974).

Braun, D. (2017). Indexicals. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017). Metaphysics Research Lab, Stanford University.

Burali-Forti, C. (1897). Una questione sui numeri transfiniti. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, *11*(1), 154-164. https://doi.org/10.1007/BF03015911

Cantini, A. (2009). Paradoxes, Self-Reference and Truth in the 20th Century. En D. Gabbay & J. Woods (Eds.), *Handbook of the History of Logic. Volume 5. Logic From Russell to Church* (pp. 875-1013). Elsevier.

Cappelen, H. & Dever, J. (2009). *The Inessential Indexical.* Oxford University Press.

Carnap, R. (1947). *Meaning and Necessity.* Chicago: University of Chicago Press.

Carnap, R. (1934/2001). *Logical Syntax of Language* (A. Smeaton, Trad.). Routledge (Publicado originalmente en 1934).

Castañeda, H. N. (1966). 'He': A Study in the Logic of Self-Consciousness. *Ratio*, *7*, 130-57.

Castañeda, H. N. (1989). The Language of Other Minds: Indicators and Quasi-Indicators. En *Thinking, Language, and Experience.* University of Minnesota Press.

Cook, R. T. (2009). Well-formed formula. En *A Dictionary of Philosophical Logic* (p. 312). Edinburgh University Press.

da Costa, N. (1963). *Sistemas Formais Inconsistentes.* Editora da Universidade Federal do Paraná.

da Costa, N. (1974). On the Theory of Inconsistent Formal Systems. *Notre Dame Journal of Formal Logic*, *15*, 597-610.

Donnellan, K. S. (1966). Reference and Definite Descriptions. *The Philosophical Review*, *75*(3), 281-304.

Dummett, M. (1973). *Frege: Philosophy of language.* Harper & Row Publishers.

Eldridge-Smith, P. & Eldridge-Smith, V. (2010). The Pinocchio Paradox. *Analysis*, *70*(2), 212-215.

Evans, G. (1982). *The Varieties of Reference* (J. McDowell, Ed.). Oxford University Press.

Ezcurdia, M. (2014). Los indéxicos y la semántica de Kaplan. En M. Ezcurdia (Ed.), *Los indéxicos y la semántica de Kaplan* (pp. 5-50). Universidad Nacional Autónoma de México - Instituto de Investigaciones Filosóficas.

Fara, D. (2001). Descriptions As Predicates. *Philosophical Studies*, *102*(1), 1-42 (La autora posteriormente tomaría el apellido Fara, por lo que este trabajo a veces es atribuido a Graff, D.) https://doi.org/10.1023/A:1010379409594

Fara, D. (2016). Further Steps Towards a Theory of Descriptions as Predicates. *Analytic Philosophy*, *57*(2), 91-109. https://doi.org/10.1111/phib.12076

Field, H. (2008). *Saving Truth From Paradox.* Oxford University Press.

Forbes, G. (2003). Indexicals. En D. Gabbay & F. Guenther (Eds.), *Handbook of Philosophical Logic* (pp. 101-134). Dordrecht: Kluwer Academic Publishers.

Frege, G. (1918/2016). El pensamiento: una investigación lógica (C. U. Moulines, Trad.). En M. M. Valdés (Ed.), *Escritos sobre lógica, semántica y filosofía de las matemáticas* (pp. 321-348). Universidad Nacional Autónoma de México - Instituto de Investigaciones Filosóficas. (Publicado originalmente en 1918 como Der Gedanke. Eine logische Untersuchung Beiträge zur Philosophie des deutschen Idealismus II, 58-77).

Frege, G. (1892/2016). Sobre sentido y referencia (C. U. Moulines, Trad.). En M. M. Valdés (Ed.), *Escritos sobre lógica, semántica y filosofía de las matemáticas* (pp. 249-275). Universidad Nacional Autónoma de México - Instituto de Investigaciones Filosóficas. (Publicado originalmente en 1892 como Über Sinn und Bedeutung, Zeitschrift für Philosophie und philosophische Kritik, 100, 25-50).

Gaifman, H. (2006). Naming and Diagonalization, from Cantor to Gödel to Kleene. *Logic Journal of the IGPL*, *14*(5), 709-728. https://doi.org/10.1093/jigpal/jzl006

Gödel, K. (1944). Russell's Mathematical Logic. En S. Feferman, J. Dawson & S. Kleene (Eds.), *Journal of Symbolic Logic* (pp. 119-141). United States of America: Northwestern University Press.

Gödel, K. (1931/1986). On formally undecidable propositions of Principia mathematica and related Systems I (J. van Heijenoort, Trad.). En S. Feferman, J. W. Dawson, S. C. Kleene, G. H. Moore, R. M. Solovay & J. van Heijenoort (Eds.), *Kurt Gödel - Collected Works. Volume I. Publications 1929-1936* (pp. 145-195). Oxford University Press & Clarendon Press (Publicado originalmente como Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, Monatshefte für Mathematik Physik, 38: 173–198).

Grelling, K. & Nelson, L. (1907). Bemerkungen Zu den Paradoxien von Russell Und Burali-Forti. *Abhandlungen Der Fries'schen Schule (Neue Serie)*, *2*, 300-334.

Gutiérrez Rodríguez, E. (Ed.). (2010). Los demostrativos. En *Nueva gramática de la lengua española - Manual*. Real Academia Española - Asociación de Academias de la Lengua Española.

Halbach, V. (2016). The Root of Evil: A Self-Referential Play in One Act. En J. van Eijck, R. Iemhoff & J. J. Joosten (Eds.), *Liber Amicorum Alberti: A Tribute to Albert Visser* (pp. 155-163). College Publications London.

Halbach, V. & Visser, A. (2014a). Self-reference in arithmetic I. *The Review of Symbolic Logic*, *7*(4), 671-691. https://doi.org/10.1017/S1755020314000288

Halbach, V. & Visser, A. (2014b). Self-reference in arithmetic II. *The Review of Symbolic Logic*, *7*(4), 692-712. https://doi.org/10.1017/S175502031400029X

Heck, R. (2007). Self-Reference and the Languages of Arithmetic. *Philosophia Mathematica*, *15*(1), 1-29. https://doi.org/10.1093/philmat/nkl028

Hintikka, J. (1975). Carnap's Heritage in Logical Semantics. En J. Hintikka (Ed.), *Rudolf Carnap, Logical Empiricist: Materials and Perspectives* (pp. 217-242). D. Reidel Pub. Co.

Hofstadter, D. R. (1979/1999). *Gödel, Escher, Bach: and Eternal Golden Braid* (2ª ed.). Basic Books.

ISO. (2009). *Quantities and units - Part 2: Mathematical signs and symbols to be used in the natural sciences and technology (ISOO 80000-2:2009)*. International Organization for Standardization.

Kant, I. (1781/2016). *Crítica de la Razón Pura* (4ta. ed). España: Taurus (Publicado originalmente en 1781 por Johann Friedrich Hartknoch (ed.)

Kaplan, D. & Montague, R. (1960). A Paradox Regained. *Notre Dame Journal of Formal Logic*, *1*(3), 79-90. https://doi.org/10.1305/ndjfl/1093956549

Kaplan, D. (1978). Dthat. En P. Cole (Ed.), *Syntax and Semantics* (pp. 221-243). Academic Press.

Kaplan, D. (1989a). Afterthoughts. En J. Almog, J. Perry & H. Wettstein (Eds.), *Themes From Kaplan* (pp. 565-614). Oxford: Oxford University Press.

Kaplan, D. (1989b). Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals. En J. Almog, J. Perry & H. Wettstein (Eds.), *Themes From Kaplan* (pp. 481-563). Oxford: Oxford University Press.

Kleene, S. C. (1952). *Introduction to Metamathematics* (N. G. de Bruijn, J. de Groot & A. C. Zaanen, Eds.; Vol. 1). Wolters-Noordhoff Publishing; North-Holland Publishing Company.

Kleene, S. C. (1986). Introductory note to 1930b, 1931 and 1932b. En S. Feferman, J. W. Dawson, S. C. Kleene, G. H. Moore, R. M. Solovay & J. van Heijenoort (Eds.), *Kurt Gödel - Collected Works. Volume I. Publications 1929-1936* (pp. 126-141). Oxford University Press & Clarendon Press.

König, J. (1905). Uber die Grundlagen der Mengenlehre und das Kontinuumsproblem. *Mathematische Annalen*, *61*, 156-160.

Kripke, S. (1975). Outline of a Theory of Truth. *Journal of Philosophy*, *72*(19), 690-716. https://doi.org/10.2307/2024634

Landini, G. (2009). Russell's Schema, Not Priest's Inclosure. *History and Philosophy of Logic*, *30*(2), 105-139.

Leitgeb, H. (2002). What is a Self-Referential Sentence? Critical Remarks on the Alleged (Non-)Circularity of Yablo's Paradox. *Logique and Analyse*, *177*(178), 3-14.

Levi, B. (1908). Antinomie Logische? *Annali di Matematica (terza serie)*, *15*, 188-216.

Lewis, D. (1979). Attitudes De Dicto and De Se. *Philosophical Review*, *87*, 513-545.

Linsky, B. & Zalta, E. N. (1995). Naturalized Platonism Versus Platonized Naturalism. *Journal of Philosophy*, *92*(10), 525-555. https://doi.org/10.2307/2940786

Linsky, B. & Zalta, E. N. (2019). Mathematical Descriptions. *Philosophical Studies*, *176*(2), 473-481. https://doi.org/10.1007/s11098-017-1024-0

Linsky, L. (1983). *Oblique contexts*. The University of Chicago Press.

Luján, M. (1999). Expresión y omisión del pronombre personal. En I. Bosque & V. Demonte (Eds.), *Gramática Descriptiva de la Lengua Española. Sintaxis básica de las clases de palabras* (pp. 1275-1315). Espasa - Real Academia Española.

Lycan, W. G. (1994). The Trouble with Possible Worlds. En *Modality and Meaning* (pp. 3-24). Dordrecht, Springer Netherlands. https://doi.org/10.1007/978-94-011-0936-9_1

Magidor, O. (2015). The Myth of the De Se. *Philosophical Perspectives*, *29*, 249-283.

Martin, C. J. (1986). William's Machine. *The Journal of Philosophy*, *83*(10), 564-572. https://doi.org/http://www.jstor.org/stable/2026432

Merleau-Ponty, M. (1945/1994). *Fenomenología de la Percepción* (J. Cabanes, Trad.). Planeta-Agostini (Publicado originalmente en 1945 como *Phénoménologie de la perception*, por Éditions Gallimard).

Milne, P. (2007). On Gödel Sentences and What They Say. *Philosophia Mathematica*, *3*(15), 193-226. https://doi.org/https://doi.org/10.1093/philmat/nkm015

Mirimanoff, D. (1917). Les antinomies de Russell et de Burali-Forti et le problème fondamental de la théorie des ensembles. *Enseignement mathématique*, *19*(1-2), 37-52. https://archive-ouverte.unige.ch/unige:113777

Montague, R. (1962). Theories Incomparable with Respect to Relative Interpretability. *Journal of Symbolic Logic*, *27*(2), 195-211. https://doi.org/10.2307/2964114

Mostowski, A., Robinson, R. M. & Tarski, A. (1953). Undecidability and essential undecidability in arithmetic. En *Undecidable Theories* (pp. 37-74). Elsevier Science - North-Holland Publishing Company.

Ninan, D. (2016). What is the Problem of De Se Attitudes? En M. García-Carpintero & S. Torre (Eds.), *About Oneself: De Se Thought and Communication* (pp. 86-120). Oxford University Press.

O'Connor, J. & Robertson, E. (2005). Philip Edward Bertrand Jourdain. En *MacTutor History of Mathematics*. School of Mathematics; Statistics University of St Andrews, Scotland. https://mathshistory.st-andrews.ac.uk/Biographies/Jourdain/

Parsons, D. (2016). *Theories of Intensionality. A Critical Survey*. Singapore: Springer.

Perry, J. (1993a). Frege on Demonstratives. En *The Problem of the Essential Indexical and Other Essays*. Oxford University Press.

Perry, J. (1993b). The Problem of the Essential Indexical. En *The Problem of the Essential Indexical and Other Essays* (pp. 33-50). Oxford University Press.

Perry, J. (1997). Indexicals and Demonstratives. En B. Hale & C. Wright (Eds.), *A Companion to the Philosophy of Language* (pp. 486-612). New Jersey: Blackwell.

Perry, J. (2009). Directing Intentions. En J. Almog & P. Leonardi (Eds.), *The Philosophy of David Kaplan* (pp. 187-201). Oxford University Press.

Picollo, L. (2018). Reference in Arithmetic. *Review of Symbolic Logic*, *11*(3), 573-603. https://doi.org/10.1017/s1755020317000351

Picollo, L. (2020a). Alethic Reference. *Journal of Philosophical Logic*, 1-22. https://doi.org/10.1007/s10992-019-09524-w

Picollo, L. (2020b). Reference and Truth. *Journal of Philosophical Logic*, 1-36. https://doi.org/10.1007/s10992-019-09525-9

Poincaré, H. (1906). Les mathématiques et la logique. *Reviste de Métaphysique et Morale*, *14*, 294-317.

Priest, G. (1994). The Structure of the Paradoxes of Self-Reference. *Mind*, *103*(409), 25-34. https://www.jstor.org/stable/2253956

Priest, G. (2000). On the Principle of Uniform Solution: A Reply to Smith. *Mind*, *109*(433), 123-126. https://doi.org/10.1093/mind/109.433.123

Priest, G. (2006). *Doubt Truth to Be a Liar*. Oxford University Press.

Priest, G. (2007). Paraconsistency and Dialetheism. En D. Gabbay & J. Woods (Eds.), *The Many Valued and Nonmonotonic Turn in Logic* (pp. 129-204). North Holland. https://doi.org/doi.org/10.1016/S1874-5857(07)80006-9

Priest, G. (2010). Inclosures, Vagueness, and Self-Reference. *Notre Dame Journal of Formal Logic*, *51*(1), 69-84. https://doi.org/10.1215/00294527-2010-005

Priest, G. (2012). Vague Inclosures. En K. Tanaka, F. Berto, E. Mares & F. Paoli (Eds.), *Paraconsistency: Logic and Applications* (pp. 367-377). Springer. https://doi.org/10.1007/978-94-007-4438-7_20

Priest, G. (2014). *One: Being an Investigation into the Unity of Reality and of its Parts, including the Singular Object which is Nothingness*. Oxford University Press.

Priest, G. (2018). *The Fifth Corner of Four: an Essay on Buddhist Metaphysics and the Catuṣkoṭi*. Oxford University Press.

Priest, G. (2019). It Was so Revolting I Couldn't Take My Eyes Off It. En A. Rieger & G. Young (Eds.), *Dialetheism and its Applications* (pp. 47-56). Cham, Springer International Publishing. https://doi.org/10.1007/978-3-030-30221-4_3

Priest, G. (1995/2002). *Beyond the Limits of Thought* (2ª ed.). Oxford University Press (Publicado originalmente en 1995).

Priest, G. (1987/2006). *In Contradiction: A Study of the Transconsistent*. Oxford University Press (Publicado originalmente en 1987 por Martinus Nijhoff).

Priest, G. (2005/2016). *Towards Non-Being*. Oxford University Press (Publicado originalmente en 2005).

Priest, G., Berto, F. & Weber, Z. (2018). Dialetheism. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Metaphysics Research Lab, Stanford University.

Priest, G. & Routley. (1989). Introduction. En G. Priest, R. Routley & J. Norman (Eds.), *Paraconsistent Logic: Essays on the Inconsistent*. Analytica: Philosophia Verlag.

Raatikainen, P. (2018). Gödel's Incompleteness Theorems. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Metaphysics Research Lab, Stanford University.

Radulescu, A. (2018). The Difference Between Indexicals and Demonstratives. *Synthese*, *195*(7), 3173-3196. https://doi.org/10.1007/s11229-017-1367-2

Ramsey, F. (1925). The Foundations of Mathematics. En *The Foundations of Mathematics and other logical essays* (pp. 338-84). Kegan Paul, Trench, Trubner & Co., LTD.

Rang, B. & Thomas, W. (1981). Zermelo's discovery of the "Russell Paradox". *Historia Mathematica*, *8*(1), 15-22. https://doi.org/https://doi.org/10.1016/0315-0860(81)90002-1

Reichenbach, H. (1947). *Elements of Symbolic Logic*. London: Dover Publications.

Richard, J. (1905). Les Principes des Mathématiques et le Problème des Ensembles. *Revue Générale des Sciences Pures et Appliquées*, *16*, 541.

Rouilhan, P. D. (1992). Russell and the Vicious Circle Principle. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *65*(1/2), 169-182. http://www.jstor.org/stable/4320280

Russell, B. (1903). *Principles of Mathematics*. Cambridge: Cambridge University Press.

Russell, B. (1906a). Les paradoxes de la logique. *Revue de Métaphysique et de Morale*, *14*(5), 627-650. http://www.jstor.org/stable/40893400

Russell, B. (1906b). On Some Difficulties in the Theory of Transfinite Numbers and Order Types. *Proceedings of the London Mathematical Society*, *4*(14), 29-53.

Russell, B. (2016). Carta de Bertrand Russell a Gottlob Frege. En M. M. Valdés (Ed.), *Escritos sobre lógica, semántica y filosofía de las matemáticas* (pp. 575-576). Universidad Nacional Autónoma de México. Instituto de Investigaciones Filosóficas.

Schroeter, L. (2017). Two-Dimensional Semantics. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017). Metaphysics Research Lab, Stanford University.

Smullyan, R. M. (1957). Languages in which self reference is possible. *The Journal of Symbolic Logic*, *22*(1), 55-67.

Smullyan, R. M. (1984). Chameleonic Languages. *Synthese*, *60*(2), 201-224. https://doi.org/10.1007/BF00485461

Smullyan, R. M. (1992). *Gödel's Incompleteness Theorems*. Oxford University Press.

Smullyan, R. M. (1994). *Diagonalization and self-reference* (Vol. 27). Clarendon Press.

Stalnaker, R. (1981). Indexical Belief. *Synthese*, *49*, 129-151.

Taylor, K. (1998). *Truth and meaning: An introduction to the philosophy of language*. Blackwell.

Weyl, H. (1918). *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis*. Leipzig: Vertag von Veit & Comp.

Whitehead, A. N. & Russell, B. (1910/1927). *Principia Mathematica* (2ª ed., Vol. 1). Cambridge University Pres (Publicado originalmente en 1910).

Zalta, E. N. (1983). *Abstract Objects: An Introduction to Axiomatic Metaphysics*. D. Reidel.

Zalta, E. N. (2019). Gottlob Frege. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019). Metaphysics Research Lab, Stanford University.