SOBRE LAS PARADOJAS DE AUTORREFERENCIA*

JUDITH SCHOENBERG
Universidad Veracruzana

Primera Parte. Hipótesis General

1. Supongamos que las Paradojas de Autorreferencia hayan permanecido sin resolver debido a que no se haya aclarado suficientemente cómo funciona la propia autorreferencia y a cuáles falacias resulta particularmente expuesto su uso. Si así fuera, habría de ser posible formular una hipótesis acerca de cómo funciona la autorreferencia, tal que haciendo uso de ella pudiéramos describir las falacias de algunas de las paradojas y mostrar por qué en dichos casos el uso correcto de la autorreferencia no conduciría a contradicción. Sería ilusorio, claro está, esperar que una hipótesis que responde a un solo modo de emplear la autorreferencia pudiera servirnos de guía para solucionar todas las paradojas. Sin embargo, si valiéndonos de semejante hipótesis lográramos solucionar varias paradojas, esto de suyo pondría en tela de juicio la suposición de que la autorreferencia como tal es inconsistente y, a la vez, proporcionaría una base para la tesis de que la autorreferencia es un modo legítimo del discurso cuyo estudio serio en sus varios aspectos constituye una tarea de la lógica.

Con esto queda esbozado en sus rasgos más generales el proyecto de este artículo. Propongo una hipótesis acerca de cómo funciona la autorreferencia, de la cual me sirvo para desarrollar sendas soluciones de dos de las antinomias que

* Estoy en deuda con Gilbert Ryle quien leyó una versión preliminar de este trabajo, y con mi colega, la señorita Rosario Amieva, tanto por sus comentarios, como por su generosa ayuda en cuestiones del idioma sin la cual no hubiera sido posible escribir este artículo en español.

han desempeñado un papel clave en la justificación de las diversas teorías jerárquicas, a saber, la Antinomia de Russell y la Paradoja de Lukasiewicz.[1]

La hipótesis que propongo no es en realidad compleja, pero su exposición se complica por algunos problemas especiales. En primer lugar, su misma novedad hace preciso que se la justifique paso por paso. Además, para poder responder de antemano a la sospecha de que esta hipótesis pudiera tener un carácter *ad hoc*, he recurrido a dos paradojas que la apoyan de maneras muy distintas. Pero mientras que la Antinomia de Russell le da apoyo desde el primer paso del análisis de esta paradoja, no es así con la Paradoja de Lukasiewicz. Ésta, por los problemas de filosofía del lenguaje que involucra, requiere de un análisis preliminar para poder ser tratada bajo la hipótesis que propongo y en comparación con la Antinomia de Russell. Por estas razones he dividido la presentación de la hipótesis en dos partes, a las que llamo respectivamente "hipótesis inicial" e "hipótesis ampliada".[2] La Segunda Parte del artículo presenta la hipótesis inicial ilustrada con el caso de la Antinomia de Russell. La Tercera Parte consiste en el análisis preliminar de la Paradoja de Lukasiewicz que desemboca en un apoyo a la hipótesis inicial. La Cuarta Parte intenta completar las soluciones de las dos paradojas mediante la hipótesis ampliada.

La hipótesis inicial es la siguiente: Algunos usos autorreferentes de las formas del discurso son particularmente aptos para dar ocasión al inesperado surgimiento, en el curso de una línea de razonamiento, de objetos hasta entonces desconocidos, o simplemente de objetos que no se encuentran ordinariamente. La clase de las clases que no son miembros

[1] Llamo "Paradoja de Lukasiewicz" a la antinomia de éste que fue hecha famosa por Tarski bajo el nombre del "Mentiroso", porque distingo entre esta paradoja y las que involucran el concepto "mentir". Véase mi artículo, "Belief and Intention in the Epimenides", *Philos, and Phen. Research,* Dic. 1969.

[2] Esta división de la hipótesis no es arbitraria. Por el contrario, corresponde a los dos pasos principales del análisis del problema: primero, el de reconocer el fenómeno lógico que está en cuestión; y, segundo, el de averiguar la explicación de ese fenómeno.

de sí mismas, descubierta por la conocida línea de razonamiento de Russell, puede considerarse como un caso paradigmático de un objeto tal, hasta entonces desconocido. Ahora bien, el problema peculiar planteado por tales objetos es que no siempre resultan aplicables a ellos los patrones de inferencia que hemos venido asumiendo. La falacia de una paradoja de autorreferencia puede que se insinúe justamente en el punto donde se recurre a extrapolar un acostumbrado patrón de inferencia para aplicarlo a un objeto, inesperadamente surgido, para el cual no rige dicho patrón. Así, desentrañar la falacia de semejante paradoja consistiría en determinar por qué cierto patrón de inferencia no se aplica a cierto objeto. Además, sería preciso determinar si hay o no un patrón de inferencia que se ajuste al carácter del objeto con resultados lógicamente aceptables.

Se presenta entonces el problema de la estructura lógica del surgimiento de dichos objetos y de la justificación del uso de esa estructura en cualquier caso dado. A este problema responde la hipótesis ampliada.

Partimos de la consideración de que los usos de la autorreferencia son característicamente usos de formas lingüísticas lógicamente contingentes, de modo que requieren ser justificados de una u otra manera. Por ejemplo, si un uso de la autorreferencia consiste en dar a una variable de una función cierto valor que plantea un caso autorreferente, la justificación de este uso de la autorreferencia viene a ser una cuestión acerca de la razón que permita asignar dicho valor a la variable. Ahora bien, surge una dificultad especial a este respecto, porque el valor que plantea el caso autorreferente puede pertenecer a una subclase inesperada de la variable, es decir, a una subclase que no se tiene en cuenta en la definición de variable que implícitamente se está asumiendo. Éste es precisamente el mecanismo lógico del surgimiento de objetos extraños o inesperados en situaciones autorreferentes, y el uso de este mecanismo está expuesto a falacias peculiares. Puede suceder que la equivocación llegue al gra-

do de admitir un valor que ni siquiera plantea de modo legítimo un caso autorreferente, como ocurre en la Paradoja de Lukasiewicz.[3] Y, sea o no equivocado el valor, puede suceder que no se reconozca el sentido preciso en que éste es extraño, como ocurre tanto en la Paradoja de Lukasiewicz como en la Antinomia de Russell. Así, en ambas eventualidades, puede insinuarse la falacia de extrapolar al valor considerado un patrón de inferencia que no le conviene.

Resta por decir aquí que el análisis que hago de la Antinomia de Russell conduce a un resultado sorprendente. Dada la validez de las premisas de la paradoja, y puesto que la contradicción desaparece, debemos afrontar el problema de caracterizar los extraños objetos lógicos que han surgido en el curso del razonamiento. Por la analogía que se pone de manifiesto entre éstos y los números imaginarios, me atrevo a sugerir que se consideren como objetos lógicos imaginarios. En cuanto a la Paradoja de Lukasiewicz, se sostiene que el objeto lingüístico surgido —una oración viciosamente circular— es en sentido estricto defectuoso en cuanto a su forma y que, en la ausencia de premisas válidas que la pudiesen justificar, es inadmisible. Sin embargo, si recurrimos a la hipótesis ampliada, no resulta difícil vislumbrar una solución al problema de la interpretación de la oración singular autorreferente.

Segunda Parte. Análisis Preliminar de la Antinomia de Russell

2. Lo impresionante acerca de la Antinomia de Russell es que una línea de razonamiento perfectamente directa e inexorable, que maneja principios lógicos universalmente aceptados y de los más fundamentales, conduce rápidamente al surgimiento de objetos lógicos sumamente extraños que además parecen involucrar contradicción. En este respecto, el

[3] Se considera aquí que una oración del tipo utilizado por Lukasiewicz para construir su oración autorreferente 's' es, desde el punto de vista lógico, un tipo muy especial de forma proposicional, o sea, un patrón verbal de significado perteneciente a un idioma particular. Véase la nota 16.

problema que la Antinomia plantea a la lógica es enteramente análogo al que le fue planteado a las matemáticas por el surgimiento, como consecuencia de las operaciones fundamentales de la aritmética, de los números negativos, irracionales e imaginarios. ¿No es razonable, entonces, esperar que la contradicción de la Antinomia al fin resulta ilusoria y que esos objetos descubiertos por Russell contribuyan a enriquecer a la lógica? El análisis que se hace a continuación está condicionado por este enfoque del problema.

El argumento de Russell parte de la consideración indudable de que hay clases, como la clase de las clases que tienen más de cinco miembros, que satisfacen sus propias condiciones de membrecía; y también hay clases, como la clase de los hombres, que no satisfacen sus propias condiciones de membrecía. Luego, por el principio según el cual satisfacer la condición de membrecía de una clase es ser miembro de esa clase, se sigue que hay una relación reflexiva de membrecía y que podemos hablar de una clase como miembro de sí misma o como no-miembro de sí misma. De este modo empieza Russell a descubrir un notable conjunto de objetos lógicos: primero, la propia relación reflexiva de membrecía (de clase); en seguida, la clase de las clases que son miembros de sí mismas y la clase de las clases que no son miembros de sí mismas. Llamemos w a esta última clase, como lo hace Russell en un pasaje que citaré en breve.

Ahora bien. La clase w, ¿es miembro de sí misma? Ésta es la pregunta cuya respuesta parece ser una antinomia; nos interesa, pues, examinar el patrón de inferencia que conduce a esa contradicción. Podemos mostrarlo mediante un ejemplo. Decimos,

(1) "Murmullo" es una palabra onomatopéyica.

Luego, de (1) inferimos analíticamente, por así decirlo, que la palabra "murmullo" suena como un murmullo. Esta inferencia es válida porque una palabra onomatopéyica *se define* como una palabra cuyo sonido es característico del referente de la palabra. Así tenemos,

(2) Si "murmullo" es onomatopéyica, entonces "murmullo" suena como un murmullo.

Además, es fácil ver que esta implicación es recíproca.

De manera análoga, en el caso de la pregunta de Russell acerca de *w*, se hace la suposición,

(3) *w* es miembro de sí misma.

Luego, puesto que una clase que es miembro de sí misma se define como una clase que posee la propiedad de su propia condición de membrecía, y dado que la condición de membrecía de *w* es *no ser miembro de sí misma*, tenemos,

(4) Si *w* es miembro de sí misma, entonces *w* no es miembro de sí misma.

Y, una vez más, la implicación es claramente recíproca. Así obtenemos la antinomia,

(5) *w* es miembro de sí misma si, y sólo si, *w* no es miembro de sí misma.

Este resultado, claro está, manifiesta la falla, en el caso de *w*, de un patrón de razonamiento cuya validez para todas las clases parece que se justifica por la propia noción intuitiva de clase. Sin embargo, esto de suyo no establece que no pueda aceptarse *w* como una clase. En principio es posible que *w* tenga propiedades peculiares tales que (3) engendre otro patrón de inferencia que no conduzca a contradicción y que, por consiguiente, el propio concepto de clase pueda ampliarse para incluir la subclase para la cual es válido este otro patrón. Si tenemos presente la historia del concepto de número, esta alternativa no suena improbable en lo más mínimo.

Consideremos, entonces, las propiedades peculiares de *w* que estarían en juego en este asunto. De modo preliminar notamos que:

(i) La condición de membrecía de *w* es *siempre* una fun-

ción de la condición de membrecía de una clase-miembro.[4] Por su propia naturaleza, esta condición de membrecía es una función *en el caso de* . . . Así, si entrara *w* en la relación reflexiva de membrecía, entonces, en cuanto clase-miembro, su *propia* condición de membrecía tendría que ser dada para el caso de sí misma y podría darse únicamente por la misma función.

(ii) De suyo, la condición de membrecía de *w* contiene la noción de la condición de membrecía de una clase únicamente en el papel de una variable. Por lo tanto, resulta que en el caso autorreferente la expresión "condición de membrecía" no tiene otro sentido que el sentido general de esta expresión. No hay nada análogo a la intensión de "murmullo" en el ejemplo anteriormente citado.

Que estas propiedades de *w* tendrán consecuencias inesperadas se sugiere de modo dramático por el siguiente argumento: Sea *v* la clase de las clases que no satisfacen sus propias condiciones de membrecía. Entonces, la condición de membrecía de *v* para cualquier clase, *incluso para ella misma*, es que esta clase no satisfaga su propia condición de membrecía. Y, siendo la condición de membrecía de *v*, *no satisfacer su propia condición de membrecía*, resulta: que *v* satisfaga la condición de membrecía de *v* es que *v* no-no satisfaga su propia (de *v*) condición de membrecía.

Este argumento en intensión es muy fuerte, no obstante, no desemboca en contradicción sino en tautología. Contrastémoslo ahora con el bien conocido argumento en extensión de Russell:

> "Let *w* be the class of all those classes which are not members of themselves. Then whatever class *x* may be, '*x* is a *w*' is equivalent to '*x* is not an *x*'. Hence giving to *x* the value *w*, '*w* is a *w*' is equivalent to '*w* is not a *w*'."[5]

[4] Utilizo las expresiones "clase-miembro" y "propiedad-condición de membrecía" como equivalentes respectivamente a las locuciones inglesas *"member class"* y *"membership condition property"*, reconociendo el carácter provisional de esta terminología.

[5] *Principia Mathematica*, Introduction, Ch. II, Sec. VIII.

¿Cómo podemos elegir entre estos dos argumentos cuyos resultados son tan inconsecuentes entre sí? He optado por el método de reformular la función de Russell de manera que pueda leerse tanto en intensión como en extensión.

Consideremos las funciones definitorias,

$$(\alpha)\ \mathrm{Wx} \equiv {\sim}\mathrm{Xx};\ \text{y},\ (\beta)\ {\sim}\mathrm{Wx} \equiv \mathrm{Xx}$$

siendo x la variable individual de clase y X la correspondiente propiedad-condición de membrecía. Así la mayúscula "W" es la condición de membrecía de w, y siendo W el *definiendum* —en el contexto de Wx, claro está— le asignamos la lectura "la condición de membrecía de w", a secas.[6] Nótese también que hemos suprimido en β la implícita doble negación del término a la derecha.

Ahora bien, tanto en α como en β, ¿qué valor toma X cuando x es w? Como ya se ha advertido en el punto (i), la condición de membrecía de w para cualquier clase, incluso ella misma, se da por la función ${\sim}Xx$, entonces para el caso de w esta función da ${\sim}W$ como la condición de membrecía para sí misma. Por supuesto, esto siempre ha sido reconocido. Sin embargo, lo que no se ha advertido es que ${\sim}W$ es la condición de membrecía de w para sí misma, en cuanto término clase-miembro de la relación reflexiva de membrecía. ¡Así que *${\sim}W$ es el valor que tenemos que dar a X!* Vemos ahora que es la función Xx la que adquiere el valor ${\sim}Ww$.[7] Pero Xx no es equivalente por definición a Wx, sino a ${\sim}Wx$. En efecto, ${\sim}Ww$ sirve sin más para instanciar el término a la derecha de β. Mas en α, todavía queda por negar ${\sim}Ww$. Resultan, pues, las tautologías,

[6] Podría considerarse, quizás, que hay cierta falta de rigor en tratar las equivalencias α y β como definiciones, pero he preferido conservar la forma de equivalencia de la función de Russell. En realidad, me parece que a este respecto estoy simplemente haciendo más explícito lo que está implícito en el procedimiento abreviado adoptado por el mismo Russell.

[7] En cuanto al que en los dos pasos de la instanciación se dan a X los valores diferentes W y ${\sim}W$, lo que pasa es simplemente que la forma negativa de ${\sim}Xx$ determina que en el caso autorreferente el valor de X en α y β sea negativo. Veremos en breve cómo sucede esto.

(α') Ww = ~~Ww; y, (β') ~Ww = ~Ww.

De esta manera la instanciación de α y β en el caso de w se lleva a cabo sin que surja ninguna contradicción.

Podría objetarse que aun podemos aplicar a $\sim\sim Ww$ y $\sim Ww$ el patrón de (3) a (5), de modo que la instanciación hecha aquí, lejos de eliminar la contradicción, confirma además que estas proposiciones son inherentemente ambiguas. Pero esta objección no es sostenible, porque los términos a la derecha de α y β sí *son* las definiciones de Wx y $\sim Wx$ en una forma que puede responder a la estructura lógica del caso autorreferente.

Enfoquemos este problema de otra manera, en torno al significado de "W". En β, por ejemplo, Xx es el definiens de $\sim Wx$ cuya instanciación para algún valor de x esperamos que nos dé el significado, o sea, la intensión de "W" en el caso de dicho valor; mas en el caso de w tal despliegue, por así decirlo, del significado de "W" resulta imposible bajo las condiciones de que la instanciación conserve la univocidad de "W" y maneje válidamente la negación. ¿Por qué?

Cuando, para empezar, instanciamos $\sim Xx$ para averiguar la condición de membrecía de w para sí misma, damos el valor W a X conservando la lectura asignada a "W", o sea, "la condición de membrecía de w". Esto nos da $\sim W$ como la condición de membrecía de w para sí misma. Ahora bien, $\sim W$ parece corresponder a nuestra noción intuitiva de lo que sería la condición de membrecía de w para sí misma. Pero surge la dificultad de que, si no caemos en la falacia lógica de confundir la negación de "W" con su intensión tenemos que aceptar que en la lectura de "W", el único sentido de la expresión "condición de membrecía" es su sentido general.

Con esto llegamos a la médula del asunto. ¿Es absurdo considerar que "condición de membrecía" tenga aquí únicamente sentido general? Ciertamente no lo es. Podemos reconocer que la instanciación pone a descubierto la propiedad extraña de w advertida en el punto (ii), que consiste precisa-

mente en que en "W" la expresión "condición de membrecía" tiene solamente sentido general y funciona lógicamente como una variable. Así, cuando X toma el valor W, lo único que se añade al significado de X es que se trata de la condición de membrecía *de* w.

Por supuesto que suena rara esta interpretación. Pero no debíamos esperar una solución simple de la Antinomia de Russell.

Según esta interpretación, Ww y $\sim Ww$ tienen propiedades extrañas. En cuanto par, tienen la forma de proposiciones contradictorias, mas no poseen el suficiente significado para ser verdaderas o falsas y guardar entre sí una relación lógica de función de verdad. Significan respectivamente que w satisface y que w no satisface la condición de membrecía de w, en el sentido general de "condición de membrecía". *Esto es todo lo que significan.* No obstante, no son sin-sentidos; si además no conducen a contradicción, todo el peso de la validez del argumento de Russell, que condujo al descubrimiento de la clase w, entra en juego para apoyar la aceptación de estas proposiciones como objetos lógicos legítimos.

3. Pero el análisis anterior, aunque sea sostenible, no resuelve el problema planteado por la Antinomia. Si en el caso w las funciones α y β no dan valores contradictorios de $\sim X\hat{x}$ y $X\hat{x}$, esto parece constituir de suyo una falla de las mismas funciones α y β. Este resultado no puede ser aceptado a menos que se explique como una consecuencia del carácter de α y β, en tanto que son funciones que integran como componentes constantes la relación reflexiva de membrecía y la extraña condición de membrecía W. Es claro, entonces, que no podemos seguir adelante sin adoptar alguna hipótesis acerca del conjunto de objetos lógicos extraordinarios que hemos visto surgir.

No es mi propósito aquí abogar en favor de la hipótesis a este respecto que presento en lo que sigue. La presento como a primera vista plausible y en cierto modo natural. Mi

sugerencia es que estos objetos, por la analogía que parece haber entre ellos y los números imaginarios, pueden considerarse como objetos lógicos imaginarios. No intento elaborar dicha analogía, cosa que yo no podría hacer puesto que no soy matemático. Me limito a considerar algunas propiedades lógicas de estos objetos por las cuales la analogía me parece brotar de suyo, por así decirlo.

Partimos de la suposición de que los objetos lógicos son o reales o imaginarios o . . . . . Es decir, dejamos abierta la posibilidad de objetos mixtos. Asumimos que la relación reflexiva de membrecía (de clase), la clase *w* y otras clases semejantes, son imaginarias. En cuanto a proposiciones, podríamos decir que enunciados como $\hat{w}w$, que a su modo peculiar carecen de suficiente significado para ser verdaderas o falsas, son imaginarias. Así, la noción de la membrecía reflexiva de *w* da origen a proposiciones imaginarias. Diríamos que tales proposiciones son inequívocamente imaginarias.

Por otra parte, también hay enunciados como,

(6) La clase de las clases que tienen más de cinco miembros es miembro de sí misma.

Puede ahora impresionarnos que (6), que seguramente encierra alguna verdad, es un compuesto de elementos reales e imaginarios y, además, tiene dos interpretaciones diferentes pero interrelacionadas. Para ver cómo es esto, nos ayudará un simbolismo sencillo. Sea *b* la clase de las clases que tienen más de cinco miembros, *u* la clase de las clases que son miembros de sí mismas, y usemos "i" y "e" suscritos a los símbolos proposicionales para indicar respectivamente que la lectura es intensional o extensional. Entonces, (6) dice a la vez: (P_e) *b* es miembro de *b*, y (Q_i) *b* posee la propiedad de ser miembro de sí misma. Luego, (P_e) tiene la lectura intensional correspondiente: (P_i) *b* es una clase que tiene más de cinco miembros; mientras que (Q_i) tiene la lectura extensional correspondiente: (Q_e) *b* es miembro de *u*. Ahora

bien, de estas cuatro lecturas (P_i) se enuncia totalmente en el dominio de objetos lógicos reales y es inequívocamente verdadera. Pero el resto de los miembros de este conjunto relacionan, cada uno a su manera, un objeto lógico real, la clase *b*, con un objeto lógico imaginario, ya sea como término de una relación imaginaria (en P_e), o como poseyendo una propiedad imaginaria (en Q_i), o como perteneciendo a una clase imaginaria (en Q_e). Entretanto, los cuatro miembros de este conjunto se relacionan todos entre sí por el mismo patrón de inferencia que hemos visto que falla para las proposiciones enteramente imaginarias, $\mathcal{W}w$ y $\sim\mathcal{W}w$. Es así muy tentador pensar que el patrón no falla en el caso del conjunto implicado por (6), debido a que este conjunto contiene la real (P_i) en la cual los objetos imaginarios desaparecen del razonamiento.

Veremos en la Cuarta Parte cómo este modo de abordar el problema puede extenderse a las funciones α y β, con el resultado de que la Antinomia se resuelva completamente.

Tercera Parte. Análisis Preliminar de la Paradoja de Lukasiewicz

4. La Paradoja de Lukasiewicz explota la forma semántica que Tarski llama la forma (T)[8] y se formula de la siguiente manera:

(T) X es verdadera si, y sólo si, p,

siendo p una oración del lenguaje al que pertenece también "verdadero", y siendo X el nombre de p.

Consideremos la paradoja aproximadamente en la forma

[8] Para evitar confusión conservo la "T" del nombre en inglés, "the form (T)"; pero debe entenderse que la "T" se refiere a "Truth", o "Verdad". Véase Alfred Tarski, "The Semantical Conception of Truth", Sec. 4. Curiosamente, parece haber sido pasado por alto que esta forma, que Tarski atribuye a Lesniewski, fue utilizada por el propio Aristóteles en las *Categorías* (*12*, 14^b11 ss). Asumo aquí, como hipótesis de trabajo, la validez de esta forma dentro del lenguaje natural.

que le dió Tarski.[9] Escribimos,

La oración impresa en la pág. 125, líneas 2 y 3, de este artículo no es verdadera.

Llamamos a esta oración 's' y luego formulamos la equivalencia de la forma (T),

(I) 's' es verdadera si, y sólo si, la oración impresa en la pág. 125, líneas 2 y 3 de este artículo no es verdadera. Entretanto, descubrimos empíricamente que,

(II) "s" es idéntica a la oración impresa en la pág. 125, líneas 2 y 3 de este artículo.

Luego, puesto que (II) establece una identidad, por la ley de Leibniz, " 's' " y "la oración impresa en la pág. 125, líneas 2 y 3 de este artículo" son expresiones intercambiables. Así, sustituyendo la última por la primera en el término de la derecha de (I), obtenemos,

(III) 's' es verdadera si, y sólo si, 's' no es verdadera.

Para empezar el análisis de esta paradoja, consideremos la Prop. (II). Esta premisa, que tiene la forma de una identidad, se aduce para fundamentar la sustitución que conduce a la antinomia. Entonces, ¿cuál es el significado de la Prop. (II) con que se pretende justificar la sustitución?

Es evidente que, si la sustitución es válida, debe estar ya determinado que 's' y la oración acerca de la cual versa 's' son exactamente la misma oración. Desde luego, si (II) no significara esto, la paradoja sería un mero sofisma verbal. Pero esta identidad se establece sólo si la expresión "la oración impresa en la pág. 125, líneas 2 y 3 de este artículo" denota inequívocamente una y la misma oración en sus tres usos en el enunciado de la paradoja. Parecería asumirse que

[9] Tarski, op. cit., Sec. 7.

esta expresión tiene un sólo referente posible y, por lo tanto, debe leerse en el sentido de una descripción definida. Así, si la Prop. (II) fundamenat la sustitución, significa que hay una y sólo una oración en la pág. 125, líneas 2 y 3 de este artículo y que esta oración es 's'.[10]

Mas no es difícil ver que según esta interpretación la Prop. (II) no es una verdad empírica sino un juicio teórico muy problemático. Esto se sigue de las propias condiciones del problema. El uso de " 's' " para instanciar la forma de locución de verdad de la forma (T) establece de suyo que se está tomando a 's' como cierto enunciado. Ahora bien, Lukasiewicz nos enseñó cómo construir esta clase de enunciado poniendo en cierto lugar una oración del tipo que puede usarse para hacer enunciados diferentes. Entonces, ¿no está también allí, en la pág. 125, líneas 2 y 3 de este artículo, la oración verbal, por así llamarla, que fue utilizada para hacer el enunciado 's'? Esta cuestión, independientemente de cómo la contestemos, pertenece claramente a la filosofía del lenguaje. Además, no es de ningún modo evidente que no pueda haber allí, en dicho lugar, más de un solo enunciado hecho por el uso de la oración verbal en cuestión. Este asunto requiere ser analizado. Así, cualquier juicio acerca de qué objeto u objetos sentenciales están allí disponibles para ser denotados por el sujeto de 's', es un juicio teórico. Ahora bien, la Prop. (II), por su significado que fundamenta la sustitución, es semejante juicio teórico. Por lo tanto, podemos concluir con toda seguridad que la Prop. (II) no es empírica.

Esta consideración, claro está, no resuelve la paradoja, pero sí la pone en una nueva perspectiva. Es claro ahora que no hay un hecho empírico ineludible que bloquee el tratamiento teórico del problema de la interpretación de la oración singular autorreferente. Además, se vislumbra la posibilidad de interpretar tal oración como versando acerca de la oración verbal de la cual es un uso. Esto pone entonces la

[10] Esta interpretación la ha hecho más o menos explícita Tarski mismo en su artículo, "Truth and Proof", en *The Scientific American*, Vol. 220, Núm. 6, junio, 1969.

interpretación del círculo vicioso, la cual es implicada por la Prop. (II), en la posición de ser únicamente otra pretensa solución. De modo que el problema de la paradoja al que debemos responder primero es: ¿No hay base teórica para excluir la interpretación del círculo vicioso como un uso incorrecto del lenguaje? Se ha considerado muy difícil esta pregunta, sin embargo me parece que podemos contestar sin rodeos que sí la hay. La dificultad es que no está a la vista la razón por la cual una oración circular como la 's' de la paradoja no está bien formada. Tenemos que indagar esa razón.

Pero antes de abordar este problema, hay otro asunto que debemos considerar.

5. ¿Qué podemos decir de la interpretación semántica de esta paradoja según la cual la contradicción se debe al uso de locuciones de verdad dentro del lenguaje natural? Me parece posible mostrar de modo concluyente que esta paradoja, debido a su propia estructura lógica, no puede poner en duda la validez de la forma (T) dentro del mismo contexto lingüístico de la paradoja, y que ni siquiera es compatible la concepción de esta paradoja con considerar que 's' sea contradictoria *porque* es una locución de verdad.

Aceptemos por mor del argumento que la antinomia se deriva válidamente. Ahora bien, el argumento procede usando 's' para instanciar la forma (T) y la equivalencia resultante, la Prop. (I), se muestra que contiene la contradicción latente que llega a ser explícita en la Prop. (III). Entonces, si queremos interpretar lo que el argumento comprueba acerca de la fuente de la contradicción, tenemos que decidir si estamos usando 's' para probar la forma (T) o la forma (T) para probar 's'. Es fácil ver que la Prop. (I) es un contraejemplo de la validez de la forma (T) sólo si 's' es indudablemente consistente, o que puede inferirse que la fuente de la contradicción es 's' sólo si se asume que la forma (T) es válida.

Precisemos esto. Abreviemos " 's' es consistente" y "La forma (T) es válida" por *S* y *T* respectivamente. Entonces, lo que se comprueba por la derivación de la paradoja, si ésta es válida, es,

(Si S, entonces no-T) y (Si T, entonces no-S).

Es decir, *S* y *T* son contrarias y esto nos da como única inferencia cierta, *no-S o no-T*. Luego, la única manera de concluir que 's' es inconsistente sería decir, *T, y por lo tanto, no-S*. Ni hay razón alguna para no decir esto, puesto que sería absurdo contraponer la consistencia de 's', como ésta se interpreta por la Prop. (II), a la validez evidente de la forma (T) que, por lo que sé, a fin de cuentas no hay quien la niegue. Pero si afirmamos *T* para inferir la inconsistencia de 's', concedemos así la consistencia de la forma de locución de verdad que es un término de la forma (T), y se sigue que no podemos concluir que 's' es inconsistente *porque* es una locución de verdad.

Estas consideraciones me parece que invalidan la interpretación semántica de la paradoja. Por fuerza, debemos buscar en la autorreferencia de 's', como ésta se interpreta por la Prop. (II), la explicación de la contradicción.

6. Intuitivamente nos parece que una oración que es un círculo vicioso no está bien formada, sin embargo no es evidente cuál sea la regla de formación violada. Quisiera sugerir ahora que esa clase de oración encierra una contradicción que atañe a su propia forma, y que esta contradicción la hace inestable.[11]

Consideremos la 's' de la paradoja. En su interpretación del círculo vicioso, se pretende que 's' sea una oración que es exactamente el mismo enunciado que el enunciado acerca del

[11] Esta interpretación del círculo vicioso la propongo aquí únicamente para el caso de la oración autorreferente *singular*. No me parece que ejemplifica esta clase de circularidad la proposición universal que puede tener una referencia a sí misma como caso de la generalización.

cual dice algo. *Pero tal oración no tiene sujeto y predicado distinguibles y es enteramente impenetrable para el análisis gramatical. Seguramente es en esto justamente en lo que consiste su circularidad viciosa.* Analizar tal oración en sujeto y predicado es destruirla. Así que en este respecto, 's' no está en la forma "S es P". No obstante, la interpretación del círculo vicioso recurre a la vez al supuesto de que 's' sí está en la forma "S es P", ya que se asume que 's' tiene un término-sujeto que denota 's' y que 's' versa acerca de 's'. De esta manera la concepción de 's' involucra la contradicción flagrante de que 's' está y no está en la forma "S es P". Ahora bien, para razonar *a partir de* tal oración, se requiere que ésta sea analizada en sujeto y predicado. Entonces, tal análisis destruirá el círculo vicioso, la oración circular se desvanecerá, y aparecerá en su lugar una oración diferente.

El enunciado de la Paradoja de Lukasiewicz es un ejemplo tan claro de esto como pudiéramos desear. El término a la derecha de la Prop. (I) se presenta como la 's' circular. Luego la sustitución es una operación analítica que al aplicársele la destruye. 's' se desvanece y aparece en su lugar otra oración que puede servir como término de la antinomia.

Nótese que si (III) es una antinomia, entonces ambos términos del "si y sólo si" deben estar en la forma sujeto-predicado, y además deben versar ambos acerca de una y la misma oración. Pero el término a la izquierda de (I), "'s' es verdadera", que subsiste sin cambio en (III), no es interpretable como autorreferente, sino que versa acerca de la 's' circular. Entonces, si el término a la derecha de (III) es su contradictorio, bien podemos preguntar, ¿cómo diablos apareció allí? Esta es la pregunta estrictamente lógica que plantea la paradoja y ahora la podemos contestar. Apareció allí esa proposición contradictoria no como el resultado de una sustitución válida en 's' sino como el resultado de la disolución de 's'. La Ley de Leibniz fue usada inválidamente para explotar un aspecto de la forma ambigua de la 's' circular y así disolverla.

De esta manera descubrimos que la razón para rechazar una oración autorreferente que se interpreta como círculo vicioso no es que pueda conducir a contradicción sino que no conduce a nada. Tal oración es defectuosa en su forma y no es posible razonar a partir de ella. Por ser un intento de razonar a partir de tal oración, la Paradoja de Lukasiewicz tiene que ser calificada como un pseudo-argumento. Sin embargo, esto no resuelve el problema planteado por la paradoja.

Hemos visto surgir por un uso de la autorreferencia un objeto extraño: la oración viciosamente circular. Tal oración aparenta tener la forma sujeto-predicado, pero resulta que por no tenerla inequívocamente le es inaplicable a tal oración *cualquier* patrón de inferencia para razonar *a partir de* enunciados en dicha forma. En fin, se trata de un objeto surgido que no puede aceptarse como un objeto lingüístico legítimo. Entonces, ¿no imputa este resultado la autorreferencia como tal? ¿No es la 's' de la paradoja un uso natural de la autorreferencia, por así decirlo, que se torna en contraejemplo del carácter confiable de la propia autorreferencia? Si es así, de nada sirve para defender la tesis de que la autorreferencia sea una clase del discurso lógicamente confiable, la posibilidad de que según otra interpretación de la oración autorreferente, ésta resulte bien formada. En esta situación, el único recurso que nos queda para defender que la autorreferencia sea una clase del discurso aceptable, es postular que, lejos de ser un uso natural de la autorreferencia, la 's' de la paradoja es un uso inadmisible de ella. ¿Es sostenible este postulado? Averigüémoslo.

Cuarta Parte. El Funcionamiento de la Autorreferencia en los Problemas de Russell y Lukasiewicz

7. Hay que precisar cuál es el fundamento de la pretensión de que una paradoja pueda imputar la autorreferencia como tal. Característicamente las formas lógicas usadas por el discurso autorreferente son formas contingentes. Por ejemplo,

las funciones α y β de la Antinomia de Russell están en la forma contingente, $\varphi x \equiv \psi x$,[12] y en la Paradoja de Lukasiewicz es la forma *S es P* de la cual se pretende hacer un uso autorreferente. Ahora bien, la deducción de una contradicción a partir de una premisa en forma contingente no imputa esa forma sino que únicamente implica que se ha puesto un contenido inapropiado en esa forma. Es evidente, pues, que tales usos inapropiados de las formas lógicas serán o autorreferentes o no-autorreferentes. Sería absurdo considerar que cualquier uso autorreferente de una forma contingente que condujera a una contradicción sería un contraejemplo de la consistencia de la autorreferencia. La contradicción puede imputar la autorreferencia como tal sólo si se justifica, por una buena razón, ese uso autorreferente de la forma en cuestión. Así tendríamos un conflicto de criterios y una paradoja, siendo que alguna razón apoya la aceptación mientras que la contradicción apoya el rechazo de cierto uso autorreferente de una forma contingente.

Consideremos ahora si las paradojas bajo consideración hacen usos justificados de la autorreferencia. En ambos casos un término variable de una expresión lingüística toma cierto valor *permitido* que da lugar a un caso autorreferente. Así, la justificación para los usos respectivos de la autorreferencia es en ambos casos una cuestión acerca de la razón que permite aceptar cierto valor de una variable.

No puede haber la menor duda de que en este sentido la Antinomia de Russell hace un uso justificado de la autorreferencia. Un argumento poderoso e intuitivamente claro justifica cierta función, y esta función es ella misma construída sobre la aceptación de *w* como una clase. Entonces, es muy difícil ver cómo puede aceptarse la función y, a la vez, excluir a *w* como un valor permitido de la variable de clase. Cualquier solución de la paradoja que recurriese a tal exclu-

[12] Claro que aún si formuláramos estas funciones como *identidades* definitorias, esto no cambiaría su carácter de contingentes desde el punto de vista formal.

sión del caso de *w* tendría inevitablemente un carácter insatisfactorio.

Mas por el contrario, la Paradoja de Lukasiewicz sí puede resolverse por la exclusión del valor que plantea el caso autorreferente considerado. Se trata de una oración verbal —que desde ahora llamaré 's_v'— en la forma de locución de verdad, y es claro que el único requisito que tiene que cumplir un uso enunciativo de 's_v' es que tal enunciado sea un uso correcto del lenguaje.[13] Así que el único criterio por el cual la 's' de la paradoja podría ser justificada es precisamente ese criterio por el cual ha sido rechazada ya.[14] En este caso no hay un conflicto de criterios y, por esta razón, no hay realmente paradoja sino simplemente un uso injustificado de una forma contingente. La posibilidad *prima facie* de construir la 's' defectuosa de la paradoja no es más que un ejemplo de la posibilidad de usar el lenguaje incorrectamente. El hecho es que no hay ningún argumento, por débil que fuera, que dé origen a la 's' de la paradoja como un caso permitido de una forma proposicional. *¡'s' simplemente se pone allí, como caída del cielo, sin razón alguna!* Y puesto que es una construcción absurda, no es admisible ni como un uso de la autorreferencia ni como un uso de la forma *S es P*. ¿Cómo, entonces, pudo impresionarnos como un uso natural de la autorreferencia?

8. Me parece que el problema del concepto de autorreferencia nos proporciona un bonito ejemplo de los Ídolos del Mercado de Bacon. Se toma el significado literal y pobre de la palabra, "autorreferencia", como si fuera una descripción del conjunto de fenómenos lingüísticos al cual dicha palabra ha sido puesta como nombre. El nombre, claro está, responde

[13] Utilizo el adjetivo "enunciativo" como equivalente a la locución inglesa, "*statement-making*", empleada por Strawson. Así, un uso enunciativo de 's_v' es un enunciado.

[14] Véase la Sec. 6. En efecto, tenemos aquí un ejemplo de que un uso inapropiado de una forma no conserva las propiedades de la forma, de modo que, estrictamente hablando, no es un uso de la forma sino un intento fracasado de usar la forma.

en cierto modo preliminar al carácter aparente de esos fenómenos, sin embargo no se sigue que del significado literal de "autorreferencia" pueda deducirse analíticamente el contenido de la noción de autorreferencia lingüística. Para formar un concepto válido de esta potencia del lenguaje y descubrir así cuáles serían los modos naturales de explotarla, hay que ir a los fenómenos autorreferentes mismos y analizarlos.

Hemos visto que pueden surgir objetos inesperados en una situación lingüística que tenga aspecto autorreferente. Este es un hecho, sin embargo ese hecho no se explica por sí mismo. Desde el punto de vista lógico, ¿qué sería *un objeto inesperado*? ¿Qué querría decir el que *surgiera* tal objeto? En la *Primera Parte* he indicado de modo preliminar una hipótesis a este respecto. Consideremos ahora si esta hipótesis puede servir para aclarar los problemas involucrados por las paradojas de Russell y Lukasiewicz.[15]

Partimos de la noción aceptada de que los valores permitidos de una variable de una función o forma proposicional constituyen una clase. Si sigue de esto que el valor que plantea el caso autorreferente, si es permitido, es miembro de esa clase. Mas puede suceder que la propia variable encierre una ambigüedad sistemática que no se haya tenido en cuenta en la concepción de la función, y que sea justamente el caso autorreferente el que explota esa ambigüedad. Por ejemplo, puesto que 's_V' tiene la forma de locución de verdad, los valores permitidos de su término-sujeto, "La oración en tal y tal lugar", tienen que pertenecer a la subclase de oraciones que

[15] Quisiera sugerir aquí la posibilidad de que mi hipótesis ampliada que se elabora a continuación, sea tratable en términos de una noción lógica general de *grupo*. Por ejemplo, como una primera aproximación a tal enfoque del problema, quizás podría decirse que venimos asumiendo la concepción de una clase de objetos (p.e., clases, oraciones, etc.). en los cuales pueden hacerse ciertas operaciones lógicas (p.e., la operación de una función) sin que los resultados de éstas nos lleven fuera del universo del discurso en el cual venimos razonando acerca de aquéllos. Es muy tentador pensar que este tipo de enfoque del problema no sólo sería posible sino además sería lógicamente más fundamental. Sin embargo, las dificultades teóricas generales que se presentarían para una interpretación tal de mi hipótesis serían muy grandes, de modo que en este artículo no intento hacer tal interpretación, ni siquiera he orientado mi argumento a su posibilidad.

son verdaderas o falsas; pero sabemos que la oración singular autorreferente no puede versar acerca de sí misma en cuento enunciado. En cambio, *fijando nuestra atención en lo que está allí, en la pág. 125, líneas 2 y 3 de este artículo,* vemos que si 'sv' tuviera un predicado apropiado, el valor de su término-sujeto que daría lugar a este caso autorreferente pudiera ser la misma 'sv', o sea, una oración que pertenece a la clase de oraciones verbales.

De esta manera se hace patente que el uso de la autorreferencia se expone a una falacia peculiar que consiste en estrechar la variable de tal modo que no pueda permitir el valor que plantee el caso autorreferente y, a la vez, permitir este valor (u otro valor con la mera pretensión de que éste dé lugar a un caso autorreferente).

Ahora bien, puede o no ser que, aún habiéndose cometido esta falacia, la función o forma proposicional admita no obstante una reinterpretación tal que pueda permitir el caso autorreferente. Y, claro está, si el uso de la autorreferencia en cuestión es justificado, como lo es en la Antinomia de Russell, urge entonces encontrar tal reinterpretación. Pero en el caso de la 's' de Lukasiewicz, que no es un uso justificado de la autorreferencia, podemos simplemente concluir que 'sv' no admite un uso autorreferente. Es decir, podemos concluir que la propia noción de una locución de verdad singular autorreferente es de suyo falaz. Volveré después al problema de fundamentar esta conclusión. Pasemos ahora al problema más importante, el de la reinterpretación de las funciones que figuran en la Antinomia de Russell.

Las formulaciones α y β proporcionan dos ejemplos relacionados de la explotación por la autorreferencia de la ambigüedad sistemática de un término variable. Consideremos primero la variable x. En α y β, x se define como la variable de clase y puede tomar como valor cualquier clase. Entonces, esto no parece una definición tan estrecha de x que no pueda abarcar el caso autorreferente. Mas el problema es que tácitamente se está suponiendo una interpretación de "clase"

según la cual w no es una clase. Veamos por qué es ésto así.

Se supone que los pares de expresiones $(Wx, \sim Wx)$ y $(Xx, \sim Xx)$, son funciones contradictorias para todos los valores permitidos de x. De manera que si x es la irrestricta variable de clase, se sigue que si w es una clase, entonces w debe pertenecer a la clase de clases que, en el sentido exclusivo, son o no son miembros de sí mismas. Luego, si resulta que w no es tal clase, como lo implica mi propuesta solución de la paradoja, se sigue en este contexto conceptual que w no es una clase. Pudiera así pensarse que la solución que propongo apoya la opinión de que no hay tal clase. Sin embargo, mi solución no apoya esta opinión por la siguiente razón:

Seguramente podemos decir que la noción lógica de predicados complementarios funciona sólo con respecto a la predicación en el discurso verdadero o falso. Podemos entonces argumentar así:

> Si (A) "Miembro de sí misma" y "no-miembro de sí misma" son predicados complementarios para todas las clases, y (B) w es una clase, entonces, (C) las proposiciones Ww y $\sim Ww$ son verdaderas o falsas.

Ahora bien, por mi interpretación de Ww y $\sim Ww$, C es falsa; y esto implica a su vez que A o B o ambas son falsas. Entretanto, el argumento poderoso de Russell que conduce a la concepción de la clase w, no ha sido refutado; y si B no se pone en duda, entonces $\sim C$ de suyo refuta A en el caso de w. Con estas bases, rechazamos únicamente A y concluimos que: *w pertenece a una subclase de clases para lo cual no rige la relación reflexiva de membrecía.* Y, puesto que la razón por la cual no rige es que, siendo x un miembro de esta subclase de clases, el uso de Wx y $\sim Wx$ se saca fuera del dominio del discurso verdadero o falso, se sigue que: *La concepción de $(Wx, \sim Wx)$ y $(Xx, \sim Xx)$ como funciones que son contradictorias cuando se usan dentro del dominio del discurso verdadero o falso, no ha sido refutada.* Pero esto es todo lo que se necesita probar para establecer que el caso de w no invalida estas funciones, las cuales descansan en la

premisa de que w es una clase.

El problema que subsiste es, ¿cuál es la subclase de clases, a la que pertenece w, que siendo sus miembros valores de x, llevan el uso de estas funciones fuera del dominio del discurso verdadero o falso? Si recurriésemos a la noción, sugerida al final de la *Segunda Parte*, de que los objetos lógicos extraños que han surgido son imaginarios, esto nos proporcionaría un enfoque de las funciones α y β. Podríamos decir, por ejemplo, que cada miembro del par $(Wx, \sim Wx)$ contiene un término constante imaginario, "W", de modo que cuando x es una clase real, la proposición resultante es compleja. En tal caso, el uso de estas funciones estaría de acuerdo con su forma de contradictorias. Pero cuando x es una clase imaginaria, como lo es la propia w, el uso de estas funciones vendría a ser completamente imaginario y saldría fuera del dominio del discurso verdadero o falso.

Parecería que así reinterpretadas, α y β permiten el caso autorreferente y la paradoja se resuelve de modo satisfactorio. Claro que subsisten problemas para la elaboración de esta interpretación del conjunto de objetos lógicos descubierto por Russell; entre otras cosas se tendrá la necesidad de recurrir a un símbolo especial (o quizás a más de uno) para designar los componentes imaginarios de expresiones.

Por supuesto, la razón por la cual Ww y $\sim Ww$ no son ni verdaderas ni falsas se remonta al carácter peculiar de la condición de membrecía de w que se manifiesta cuando α y β son instanciadas para el caso de w. En el argumento precedente, he estado asumiendo el análisis ya hecho de este asunto. Mas reconsiderémoslo ahora desde el punto de vista de cómo el caso de w explota la ambigüedad sistemática de la variable X.

Hemos visto que la condición de membrecía de w se concibe como función de la condición de membrecía de una clase-miembro. Ahora bien, cuando X toma como valor alguna condición de membrecía que es una propiedad específica, puede determinarse si la clase de cosas que poseen esta pro-

piedad, posee o no esta propiedad. En tal caso, el valor de X se expresa por un concepto bajo el cual *puede* subsumirse algo. Pero no es así cuando x es w. En este caso el valor que damos a X, $\sim W$, no es una propiedad específica. Esto se pone de manifiesto cuando descubrimos que, para no confundir la intensión de "W" con su negación, tenemos que restringir la lectura de "W" a su sentido general. Además, tal lectura restringida de "W" concuerda perfectamente con la concepción de la clase w, porque *no hay ninguna propiedad unívocamente especificada, denotada por "la condición de membrecía de w"*. Aquí una vez más la analogía con los números imaginarios reclama nuestra atención. Tal como no hay una raíz cuadrada de —1, no hay una propiedad W. Se trata de la mera idea general de tal propiedad. A pesar de que sí podemos forzar el asunto acerca de la condición de membrecía de w para sí misma, no podemos hacer aparecer tal propiedad mediante un artificio lógico. Por el contrario, al forzar el asunto, salimos del dominio del discurso verdadero o falso.

De este modo nos damos cuenta de que la variable lógica de propiedad, simbolizada aquí por "X", puede tomar como valor la mera idea general de la propiedad-condición de membrecía de una clase, en distinción de un valor que es una condición de membrecía en sentido pleno. Podría decirse entonces que ésta es la sorprendente ambigüedad sistemática de la variable X, la cual es explotada en el caso autorreferente w de α y β. Además, el valor de X tendría el mismo carácter en el caso de u, la clase de las clases que son miembros de sí mismas. Por otra parte, otras paradojas como la de Grelling tienen la misma estructura lógica que la que tiene la Antinomia de Russell e involucran el mismo tipo de valor de la variable de propiedad. Por así decirlo, parecería que la capacidad de esta variable para ser empujada a un límite del sentido, se realiza mediante cierta estructura lógica la cual es común a varias paradojas cuyos argumentos tienen puntos de partida diferentes. Entonces, puesto que esa

estructura se caracteriza por la autorreferencia, sería muy natural considerar que se pone de manifiesto así una operación de la propia autorreferencia.

Estos resultados del análisis de la Antinomia de Russell me parecen justificar la siguiente conclusión: *Si aceptamos el argumento preliminar de Russell que conduce a la concepción de su función, entonces la Antinomia puede solucionarse mediante la hipótesis de que Russel descubrió una nueva clase de objetos lógicos.* Ahora bien, esta solución del problema tiene la ventaja de reconocer la validez del argumento preliminar, y ésta es una ventaja importante. La enorme fuerza de esta paradoja consiste precisamente en que mediante este argumento preliminar impecable e intuitivamente válido, se descubre un problema cuya solución claramente requiere un desarrollo de la teoría lógica. Así es que ninguna solución que recurra a un desarrollo teórico que a su vez invalida el argumento preliminar, puede apaciguar del todo la inquietud lógica a la que esta paradoja, tan brillantemente concebida, ha dado origen.

9. Hemos podido ver que el *cómo* funciona la autorreferencia en el razonamiento acerca del problema de Russell es una consecuencia del contenido conceptual de ciertas premisas, de modo que este mismo razonamiento pone a descubierto cierta estructura lógica de la autorreferencia. Por supuesto, es la validez de esas premisas la que justifica que se recurra a esa estructura autorreferente, ya que ésta por sí misma no es más que un recurso del discurso contingente. Se trata de un uso de la autorreferencia justificado por consideraciones teóricas. En cambio, el uso justificado de la oración singular autorreferente consistiría simplemente en usar el lenguaje correctamente bajo cierta condición que se impone *sin razón.*

Examinemos lo que se encuentra en las líneas 2 y 3 de la página 125 de este artículo. Vemos que el fenómeno lingüístico que se pretende construir allí tiene que cumplir la condición de que mediante una y la misma cadena de símbolos se

escriba no sólo un enunciado que versa acerca de una oración sino también la oración acerca de la cual versa este enunciado. Ahora bien, la propia construcción que impone esta curiosa condición no es, y no podría ser, puramente lingüística. La construcción tiene un aspecto empírico, pues *el propio objeto acerca del cual versa el enunciado, no obstante que sea un objeto lingüístico, una oración, es un objeto que se hace presente, ante los ojos, mediante una señal.* Así, la oración acerca de la cual versa la oración autorreferente, aparece en el papel de un objeto señalado y encierra, a su manera, la ambigüedad esencial de cualquier objeto que se puede señalar con el dedo. Además, no existe ninguna convención según la cual hubiera una interpretación usual de tal oración. La convención me dicta que el mueble en el que estoy sentada sea usado como escritorio, y así lo uso, a pesar de que a menudo pienso que serviría mejor para guardar ropa. En cambio, la oración acerca de la cual versa la oración autorreferente no tiene ningún uso convencional. Al usarla tendremos que hacer explícito cuál es el uso que estamos haciendo de ella, o nadie nos entenderá. El problema así planteado, es: ¿Qué podría decir tal oración bajo la condición de que, con las mismas palabras se diga algo acerca de esta oración? Comprendido así el asunto, podemos expresarlo de modo abreviado: ¿Qué clase de cosa puede decirse mediante la oración singular autorreferente? Este es el rompecabezas al que da lugar cualquier modo de construir semejantes oraciones, y este rompecabezas no tiene una solución única privilegiada, o sea, necesaria.

En primer término, en el momento en que enfocamos la oración autorreferente como un enunciado que versa acerca de una oración, se hace patente que tal enunciado puede versar acerca de la oración verbal usada para enunciarlo.[16] Pero

[16] Tratar el problema de la distinción entre la oración verbal y el enunciado rebasaría los límites de este artículo, sin embargo es imprescindible tomar en cuenta algunos aspectos de este problema que son relevantes para el análisos de la oración autorreferente. Por ejemplo, es evidente que en este contexto no funciona la terminología que distingue entre "oración" y "enunciado". Hay que conservar el sentido genérico de oración, así es que prefiero usar,

hay también otras posibilidades. Consideremos un ejemplo que llamaremos 'r':

> *La oración impresa en las líneas 3 y 4 de la página 140 de este artículo está bien formada.*

Es fácil ver que debido a la potencia regresiva de esta construcción, 'r' puede interpretarse como el enunciado 'r_1' que versa acerca de otro enunciado 'r_2' que a su vez versa acerca de la oración verbal 'r_v'. O, podemos interpretar 'r' simplemente como un enunciado que versa acerca de 'r_v'. Ambas interpretaciones son posibles porque el predicado "bien formado" es aplicable tanto al enunciado como a la oración verbal. Ahora bien, consideremos otro ejemplo cuya interpretación puede explotar de modo más radical el carácter de la oración autorreferente. Llamemos 't' a la siguiente oración:

> *La oración impresa en las líneas 16 y 17 de la página 140 de este artículo involucra una regresión infinita.*

Parecería que nada nos impide interpretar 't' como un enun-

provisionalmente, "oración verbal" como equivalente de la "oración" de Strawson. Por otra parte, de acuerdo con Strawson, asumo que la oración verbal tiene significado y puede usarse de modo no-enunciativo. (Véase P. F. Strawson, *Introduction to Logical Theory*, Cap. 6, III, Sec. 10.) En efecto, tales oraciones son patrones de significado, pertenecientes a un idioma. Entre sus usos no-enunciativos se destaca el de que en el aprendizaje de un idioma, llegar a entenderlo intuitivamente y adquirir dominio de él, es en gran parte el proceso de entrenar la imaginación en sus patrones oracionales de significado. Pero esta consideración no aclara del todo el papel de la oración verbal en la oración autorreferente. Aquí entra en juego también el carácter lógico de la oración verbal de ser una forma proposicional cuyo término-sujeto es una variable. Siendo esta forma proposicional un patrón verbal de significado, está presente —no meramente abstraíble— cuando se usa de modo enunciativo. En el uso enunciativo usual del lenguaje, la presencia de la oración verbal no llama la atención. Asimismo en el caso de la oración autorreferente, el no ver que la oración verbal está allí ha obscurecido la posibilidad de la interpretación de tal oración que es a la vez la más natural y la más razonable. Es fácil ver que, siendo el término-sujeto de la oración verbal una variable sin referente, si la oración autorreferente versa acerca de la oración verbal, se corta la regresión.

ciado que versa *acerca de* una oración que involucra una regresión infinita. Puesto que 't', *qua* enunciado, podría expresarse en otras palabras y puesto que el primer paso de la interpretación de una oración autorreferente consiste en especificar cuál es la oración-objeto, por así llamarla, acerca de la cual versa cierto enunciado, esta interpretación de 't' apela simplemente a la posibilidad de decir, sin caer en una regresión, que cierta oración encierra una regresión infinita.

Vemos así que la oración autorreferente puede versar acerca de una oración verbal o acerca de un enunciado o acerca de una oración que encierra regresión infinita; además, es probable que existan otras posibilidades. En esta situación solucionar nuestro rompecabezas consistiría en descubrir de cuántas maneras diferentes es posible que la oración singular autorreferente explote la ambigüedad sistemática de "oración". Resulta, así, que esta clase de oración recurre a la misma estructura lógica de la autorreferencia a la que recurre el razonamiento acerca del problema de Russell. Esto sucede porque la oración autorreferente, según el predicado que tenga, puede versar de modo significativo acerca de una u otra clase de oración.

Preguntemos ahora: ¿Es posible que una oración autorreferente tenga el predicado "verdadero"? ¿Hay tal cosa como la locución de verdad singular autorreferente? Exploremos las posibilidades interpretativas de 's'.

(a) Si 's' versara acerca de 's_v', 's' sería un error de categoría, puesto que la oración verbal no tiene valor de verdad.

(b) Si interpretamos 's' como un 's_1' que versa acerca de un enunciado 's_2', esto plantea al menos dos alternativas. Primero, si se especifica que 's_2' versa acerca de 's_v', entonces 's_2' resulta ser un error de categoría, de modo que sería difícil que la locución de verdad 's_1' versara acerca de ella. Segundo, si 's_2' es una oración que involucra una regresión infinita de enunciados, tal oración regresiva sería ambigua y aunque aparentara decir algo verdadero o falso (lo que no ocurre en este caso), no lo diría. Así que las dos alternati-

vas tienen la consecuencia de que 's_1' resulta ser un error de categoría.

Ahora bien, este análisis de 's' no comprueba con certeza que sea imposible interpretar 's' de modo tal que 's' resultara ser un uso significativo del lenguaje, sin embargo parece muy poco probable que sea posible tal interpretación. De hecho, nadie pretende que lo sea. Y puesto que el uso autorreferente justificado de 's_v' tendría que fundamentarse en la posibilidad de tal interpretación, sí podemos concluir que, por lo que sabemos, 's_v' no permite el caso autorreferente.

De esta manera la Paradoja de Lukasiewicz queda resuelta mediante el rechazo de la locución de verdad singular autorreferente como un uso sin sentido de la oración autorreferente. Pero esto no implica que falle el uso del predicado "verdadero" dentro del lenguaje natural. Claro que si hubiera una razón que condujera con necesidad al uso autorreferente de 's_v', entonces tendríamos un conflicto de criterios y una paradoja. Mas faltando tal razón, podemos concluir sencillamente que el uso autorreferente de 's_v' involucra un uso de "verdadero" que resulta ser un error de categoría. Por lo demás, no es de ningún modo una novedad el que el uso de "verdadero" se exponga a esta falacia.

Por otra parte, resulta evidente que la oración singular autorreferente en cuanto forma de utilidad expresiva, es una mera curiosidad; sin embargo, esta clase de oración tiene otra utilidad. Debido a que tal oración pone en yuxtaposición un enunciado y la oración verbal usada para enunciarlo, el análisis de los problemas planteados por la oración autorreferente puede contribuir a aclarar la distinción entre el enunciado y la oración verbal. Así vemos una vez más cómo la autorreferencia puede servir a la propia investigación lógica.

SUMMARY

Part I. General Hypothesis

1. The possibility to which this paper responds is that the Paradoxes of Self Reference have remained unresolved because it has not been sufficiently understood how self reference itself works and to what peculiar fallacies its use is consequently exposed. I propose here a general hypothesis about the workings of self reference and, in its context, seek to resolve both Russell's and Lukasiewicz's antinomies. While I do not claim that this hypothesis can explain all uses of self reference, I do claim that if it serves to resolve some of the paradoxes, then this suggests the need to regard self reference as a legitimate kind of discourse whose peculiar utility, in its various uses, requires further study.

My hypothesis is presented in two parts. The initial hypothesis is concerned with identifying the logical phenomenon to which a certain kind of use of self reference gives rise, and is illustrated by preliminary analyses of Russell's Antinomy (Part II) and of Lukasiewicz's Paradox (Part III). The amplified hypothesis is concerned with the explanation of this logical phenomenon, and is used to complete the resolutions of the two antinomies (Part IV.)

The initial hypothesis is this: The use of self reference is sometimes apt to provide the occasion for the cropping up, in the course of a line of reasoning, of hitherto unknown or simply unusual objects. Russell's class of classes which are not members of themselves may be regarded as a paradigmatic case of such an object, hitherto unknown. Now such objects are apt to pose the difficulty that certain patterns of inference we have been assuming are inapplicable to them. Thus the fallacy of a paradox of self reference may consist in the invalid extrapolation of an accustomed pattern of reasoning to an object that has cropped up. To exhibit the fallacy of such a paradox is to determine why a certain pattern of inference does not hold for a certain object. Also, of course, one must determine whether or not there is a pattern of inference that is applicable to the object and with logically acceptable results.

The amplified hypothesis is concerned with the logical structure of such cropping up of objects and with the nature of the grounds for the use of this logical mechanism in any given case. We start from the consideration that the uses of self reference are characteristically uses of logically contingent forms, so that if some given use

of self reference leads to contradiction, this cannot impute self reference unless this use of it is somehow justified. Thus if a variable of a function or propositional form takes a certain value that poses a self referent case, then the justification of this use of self reference is a matter of the grounds for permitting this value of the variable. Now a peculiar difficulty arises in this respect, because the value that poses the self referent case is apt to belong to a subclass of the values of the variable that has not been taken into account in defining the variable. This, then, is the mechanism by whose means an unexpected object can crop up, while the problem posed is to characterize the subclass to which this object belongs. This then enables us to reconsider the whole matter of whether the function can permit a self referent case.

Part II. Preliminary Analysis of Russell's Antinomy

2. We note that in mathematics the concept of number had to be progressively enriched owing to the cropping up, by the fundamental operations of arithmetic, of negative, irrational and imaginary numbers. We then admit the possibility that an analogous problem may be posed for the conception of logical objects owing to the consequences of fundamental logical principles; and we entertain the hypothesis that Russell's Antinomy, which obviously involves the cropping up strange logical objects, poses this problem.

Russell's straightforward and compelling line of reasoning, which is grounded in universally accepted logical principles, rapidly discovers some remarkable logical objects: the reflexive class membership relation, the class of classes that are not members of themselves, and the class of classes that are not members of themeselves (this latter to be called *w*, following Russell). The difficulty then is that, by an accustomed pattern of reasoning the question about whether *w* is or is not a member of itself leads to a contradiction.

We display the accustomed pattern of inference as follows: We say,

(1) "Murmur" is an onomatopeic word.

Then, since an onomatopeic word is defined as one whose sound is characteristic of the referent of the word, we have,

(2) If "murmur" is onomatopeic, then "murmur" sounds like a murmur.

And clearly, the implication is reciprocal.

Analogously, to reason about the reflexive class membership of *w*, we assume conditionally that,

(3) *w* is a member of itself.

Then since a class which is a member of itself is defined as one which possesses its own membership condition property, and since *w's* membership condition property is *not to be a member of self*, it follows that,

(4) If *w* is a member of itself, then *w* is not a member of itself.

And again, the implication is clearly reciprocal. So we get the antinomy,

(5) *w* is a member of itself if, and only if, *w* is not a member of itself.

Now the reasoning from (3) to (5) seems to be justified by our intuitive notion of a class; yet it remains possible that *w* has peculiar properties such that (3) engenders a different pattern of inference, and my suggestion is that this latter is the case.

The class *w* has two very strange properties which can be formulated in a preliminary way as follows:

(i) *w*'s membership condition is *always* a function of the membership condition of a member class. In its nature, this membership condition is a function *in the case of*.... So if *w* itself is to enter into the reflexive class membership relation, its membership condition for itself, in the role of the member class, has to be given *before* its membership condition function can take this membership condition as an argument. Yet *w*'s membership condition for itself can only be given by the same function.

(ii) Of itself, *w*'s membership condition involves the notion of the membership condition of a class only in the role of a variable. It therefore turns out that in the self referent case of *w*'s membership condition for itself, the expression "membership condition" that enters into this membership condition has no other sense than the general sense of this expression.

That (i) and (ii) may have unexpected consequences is suggested in a dramatic way be the following argument: Let *v* be the class of classes that do not satisfy their own membership conditions. Then *v*'s membership condition for any class, including itself, is that this class does not satisfy its own membership condition. But *v's* own membership condition is, *not to satisfy its own membership condition.* So for *v* to satisfy *v's* membership condition is for *v* not-not to satisfy its own (*v*'s) membership condition!

This argument in intension is strong, and leads not to contradiction but to tautology. On the other hand, Russell's well known argument in extension (cited in English in the body of the paper) leads to contradiction. How, then, shall we choose between these two arguments? Surely the indicated procedure is to reformulate Russell's function in a way that permits both intensional and exten-

sional readings, and to reconsider the instantiation problem in this context. We may put,

$$(\alpha)\ \mathrm{Wx} \equiv \sim\mathrm{Xx};\ \text{and},\ (\beta)\ \sim\mathrm{Wx} \equiv \mathrm{Xx},$$

x being the individual class variable and X *the* corresponding membership condition property variable. In this symbolism, W is w's membership condition, and since it is defined (in the context of Wx) by the right hand term of α, we assign it the non-committal reading, "w's membership condition". Notice, too, that in β we have suppressed the implicit double negation of the right hand term.

Now then, what is the value of X when x is w? As indicated above in (1), w's membership condition for *any* class is given by the function $\sim Xx$, and for the case of w this function gives $\sim W$ as w's membership condition for itself. Of course, this has always been recognized. But what has not been taken into account is that $\sim W$ is w's membership condition for itself in its role as the member class term of the reflexive class membership relation. Thus $\sim W$ is the value we have to put for X! It is, then, the function Xx which, when instantiated for the case of w, gives $\sim Ww$. But Xx is equivalent by definition not to Wx but rather to $\sim Wx$. And, indeed, $\sim Ww$ serves without more ado to instantiate the right hand term of β, while in the right hand term of α, the instantiation of Xx as $\sim Ww$ remains to be negated. So we end up with the tautologies,

$$(\alpha')\ \mathrm{Ww} \equiv \sim\sim\mathrm{Ww};\ \text{and},\ (\beta')\ \sim\mathrm{Ww} \equiv \sim\mathrm{Ww}.$$

The contradiction has disappeared.

Moreover, this result is interpretable as corroborating the point made above in (ii). When we instantiate Xx as $\sim Ww$ this corresponds to our intuitive notion of what would be *w's* membership condition for itself. But this is because we are reading "W" unambiguously as "w's membership condition". If we tried to read any more into "W", we should fall into the fallacy of confounding the negation of "W", according to its assigned reading, with the intension of "W". To avoid this fallacy, we have to accept the assigned reading of "W" as its complete reading. And counter-intuitive as this may seem, it is consistent with point (ii). Since in w's membership condition, the expression "membership condition" has only a general sense, this expression cannot acquire any additional sense in the case of w. It is the negative form of $\sim Xx$, regarded as the function that gives w's membership condition for itself in

the first step of the instantiation, that permits the instantiation of Xx in α and β to express w's membership condition for itself.

To be sure, this is all very strange. The import of the instantiation of α and β when x is w turns out to be that Ww and $\sim Ww$ mean respectively that w does and w does not satisfy w's membership condition, in the general sense of "membership condition". *This is all they mean.*

We are thus left with an exceedingly strange pair of propositions. Ww and $\sim Ww$ are neitner nonsense nor self-contradiltory. Regarded as a pair, they have the form of a pair of contradictories, yet they have no intensional reading which can serve for judging whether they are true or false and thus ground their standing in a truth functional relation to each other. *Their meaning has a deprived character.* They would seem to be a kind of propositional object we have not met up with before; and if so, they have to be added to the set of strange logical objects which crop up in the development of Russell's line of reasoning.

3. It must be conceded, however, that the preceding argument does not resolve Russell's Antinomy. By this argument, the pairs of functions, $(Wx, \sim Wx)$ and $(Xx, \sim Xx)$, are not contradictories for the case of w. This result would invalidate the conception of α and β unless it could be shown that it is a consequence of the character of the functions α and β in so far as they contain as constant components the reflexive class membership relation and the strange membership condition W. Thus in order to go on with this line of attack on the Antinomy, we must perforce adopt some hypothesis about these strange logical objects.

It seems to me that there is a striking analogy between these objects and the imaginary number, so in what follows I entertain the hypothesis that these objects are in some sense *imaginary logical objects.* The rest of this section explores one aspect of the logical problem that has arisen that suggest this analogy. In *Part IV* I shall try to show that *if* this view were adopted, α and β would be interpretable as permitting the self referent case.

If we were to make a distinction between real and imaginary logical objects, we could classify as imaginary the reflexive class membership relation, the class w and other similar classes. Propositions like Ww, whose meaning has a deprived character and which are consequently neither true nor false, would be imaginary. On this interpretation, the notion of the reflexive class membership of w gives rise to imaginary propositions. Such propositions could be thought of as *wholly* imaginary.

On the other hand, we have also met up with propositions like,

(6) The class of classes having more than five members is a member of itself.

It may now strike us that (6) can be regarded as a composite of real and imaginary elements and that, in addition, it has two different but interrelated readings. To see how this works out, let b be the class of classes having more than five members and u the class of classes that are members of themselves. Also, let "i" and "e" be used as subscripts to indicate intensional and extensional readings respectively. Then (6) says both (P_e) b is a member of b, and (Q_i) b possesses the property, member of self. Moreover, (P_e) has the corresponding reading, (P_i) b is a class having more than five members; and (Q_i) has the corresponding reading, (Q_e) b is a member of u. Of these four readings, (P_i) is stated wholly in the domain of real objects and is unequivocally true. But the rest of the members of this set each relates a real logical object, the class b, to an imaginary logical object, whether as a term of an imaginary relation (in P_e), or as possessing an imaginary property (in Q_i), or as belonging to an imaginary class (in Q_e). Now all four members of this set are interrelated by precisely the pattern of reasoning which we have seen to fail in the cases of the wholly imaginary propositions, Ww and $\sim Ww$. It is thus very tempting to think that the pattern does not fail in the case of the set implied by (6) because this set contains the real (P_i), in which imaginary objects disappear from the reasoning.

Part III. Preliminary Analysis of the Lukasiewicz Paradox.

4. It is pointed out that the form (T), which Tarski attributes to Lesniewski and which this paradox exploits ("X is true if, and only if, p", p being a sentence of the language to which "true" belongs, and X being the name of p), was used by Aristotle himself in the *Categories* (14^b11 ff.). My assumption is that this form is valid within ordinary language and that a careful analysis of the paradox will vindicate this view.

The Lukasiewicz Paradox is stated in the later form given it by Tarski in his paper, "The Semantical Conception of Truth", Sec. 7. The self referent sentential construction 's' is on page *125*, lines *2* and *3*. Prop. (I) is the equivalence of the form (T); Prop. (II) is the identity, claimed to be an empirical fact, which is presented as the ground of the substitution; and Prop. (III) is the antinomy.

My first point then is that Prop. (II) is not empirical. If the substitution is valid and Prop. (II) is its ground, then Prop. (II) must have the import that 's' and the sentence 's' is about are the same sentence. If (II) did not have this import, then the substitu-

tion would rest on a verbal sophistry. The required import is assured by reading the expression, "The sentence on page *125*, lines *2* and *3* of this paper", in the sense of a definite description. Prop. (II) then means that there is one and only one sentence in the said place, and that sentence is 's'. But on this interpretation, Prop. (II) is a problematical theoretical premise. Since 's' is used to instantiate the truth locution form of the form (T), it is unquestionably being treated as a statement. However, 's' was constructed by putting in a certain place a sentence of the type that can be used to make different statement. It is, then, a question of linguistic theory whether this verbal sentence, as I call it, is also there on page *125*, lines *2* and *3*. Moreover, it may not simply be assumed that there is only possible statement-making use of that verbal sentence in that place. This is a matter requiring analysis. It follows, therefore, that if Prop. (II) has the import needed to ground the substitution, it is a theoretical judgment about what sensential objects are available to be denoted by the subject of 's'.

This result puts the paradox in a new light. There is no hard empirical fact blocking the path to the theoretical treatment of the singular self referent sentence. Moreover, we now glimpse the possibility of interpreting such a sentence as being about the verbal sentence used to state it; and the vicious circle interpretation becomes just another claimant. Are there, then, theoretical grounds for rejecting this latter? The attempt to answer this question is postponed until after the semantical interpretation of the paradox is considered.

5. My contention is that, owing to its logical structure, the Lukasiewicz Paradox can neither place the form (T) in doubt within its linguistir context, nor can it give ground for considering that 's' is contradictory within this context because it is a truth locution.

Let us accept for the sake of the argument that the antinomy is validly derived. Clearly, the reasoning proceeds by using 's' to instantiate the form (T). If, then there is a latent contradiction in Prop. (I) which is elicited by the substitution, this contradiction can be traced either to the form (T) or to 's', but not to both. Because: Prop. (I) is a counter-example to the validity of the form (T) only if 's' is indubitably consistent; or, if the form (T) is valid, the source of the contradiction can be inferred to be 's'. Thus all that can be inferred about the source of the contradiction is that *S* (" 's' is consistent") and *T* ("The form (T) is valid") are contraries. We have, *not-S or not-T*. Then to prove that 's' is inconsistent we must say, *T, and therefore, not-S*. Nor is there any

reason not to say this since it would be absurd to pit the consistency of 's', interpreted as a vicious circle by (II), against the self evident validity of the form (T) which, to my knowledge, no one ultimately denies. But if we assert *T* in order to infer the inconsistency of 's', we have conceded the consistency of the truth locution form which is a component of the form (T), so we cannot then turn about and say that 's' is inconsistent *because* it is a truth locution. Thus even if the paradox were validly derived (which is not the case), its semantical interpretation would be groundless. We must perforce look to the self reference of 's', on its vicious circle interpretation, for the source of the trouble.

6. It is now argued that a sentence which is a vicious circle can indeed be rejected on the ground that it is not well formed. By this interpretation, the self referent sentence purports to be indentically the same statement as the statement it says something about.* Now surely, the vicious circularity of such a sentence consists just in that it has no distinguishable subject and predicate, and so is utterly impervious to grammatical analysis. To analyse such a sentence into subject and predicate is to destroy it. In this respect, therefore, it is not in the form "S is P". Yet at the same time, the conception of such a sentence assumes that it is in the form "S is P", in that its subject denotes itself. Thus the conception of the viciously circular sentence involves the contradiction that this sentence is and is not in the subject-predicate form. It would surely be surprising, then, if one could reason validly from such a sentence.

The purported line of reasoning of the Lukasiewicz Paradox can now be reinterpreted as follows: The right hand term of (I) is the viciously circular 's'. The substitution is then an analytical operation on 's' which destroys it. 's' vanishes, and the right hand term of Prop. (III) appears. Notice that if (III) is an antinomy, then both terms of the "if and only if" must be in the subject-predicate form, and both must be about precisely the same sentence. But the left hand term of (I), " 's' is true", which subsists unchanged in (III), is non-self referent and is about the circular 's'. So if the right hand term of (III) is its contradictory, where did it come from? This is the strictly logical question posed by the paradox; and we can now answer it. The right hand term of (III) resulted from a misuse of Leibniz's Law to exploit one aspect of the ambiguous form of 's' and thus dissolve 's'.

* I wish to emphasize that in this paper I am concerned only with the *singular* self referent sentence, and not with the general proposition which may have a reference to itself as a case of the generalization. It does not seem to me that the latter kind of self reference involves vicious circularity.

It follows that the reasoning of the paradox is a pseudo-argument. 's' is to be rejected, not because it leads to a contradiction, but rather because it leads nowhere. It is not a well formed sentence from which one can reason validly.

However, there is a residual problem: Are we to regard the circular 's' of the paradox as a *natural* use of self reference that constitutes a counter-example to the reliability of self reference as such? Or, may we regard 's' a misuse of self reference. In *Part IV* it is argued that there are theoretical grounds for the latter alternative.

Part. IV. The Workings of Self Reference in Russell's and Lukasiewicz's Problems

7. It may be asked, On just what grounds can a paradox be thought to impute self reference as such? The uses of self reference are characteristically uses of logically contingent forms. Now the deduction of a contradiction from a premise in a contingent form ordinarily implies only that the content which has been put into that form is inappropriate. The contradiction becomes paradoxical only if there is nonetheless a good reason for putting that content in that form. Then we have a conflict of criteria and a paradox, some *reason* supporting the acceptance and the contradiction supporting the rejection of a given use of contingent discourse.

In both Russell's Antinomy and the Lukasiewicz Paradox, the justification of the respective uses of self reference is a matter of the reason for permitting a certain value of a variable term that poses a self referent case of a function or propositional form. In Russell's problem, we have a paradigmatic example of a justified use of self reference. Russell's unanswerable preliminary argument justifies the function which rests on accepting that w is a class, so it is very hard to see how one could accept the function and at the same time rule out w as a permitted value of the class variable. But in the Lukasiewicz Paradox, we are dealing with an arbitrarily formulated verbal sentence —hereafter called 's_v'— in the form of a truth locution. Now surely, any statement-making use of 's_v', whether self referent or non-self referent, is permissible if, and only if, it is a correct use of language. Thus the very criterion by which it could be hoped to justify the 's' of the paradox is just that criterion by which this circular 's' has already been rejected. And since there is no argument of any kind which etablishes a reason for constructing 's', there simply is no conflict of criteria in this case. The *prima facie* possibility of constructing the defective 's' of the paradox is merely an instance of the *prima facie* possib-

ility of misusing language. This construction misuses both self reference and "S is P". But if this is so, how could it strike us as a *natural* use of self reference?

8. The problem of the concept of self refernce affords a very pretty example of Bacon's Idols of the Market Place. The name, "self reference", that has been given to a set of indicated linguistic phenomena does indeed respond in a preliminary way to the appearance of these phenomena. But it does not follow that the properties of self reference or its natural uses can be analytically deduced on the ground of the antecedently given meaning of the expression, "self reference". To form a valid notion of the expressive potential of this linguistic resource, we have to go to the self referent phenomena themselves and analyse them.

We have seen that the use of self reference may lead to the cropping up of unexpected objects. To consider the logical structure of this phenomenon, I now recur to the amplified hypothesis already outlined in *Part I.**

We start from the accepted notion that the permitted values of a variable constitute a class. Now a value that poses a self referent case of a function may exploit the systematic ambiguity of the variable term, in that this value may belong to a subclass of the values of the variable that has not been taken into account in defining the variable, or perhaps has never even been conceived. If, then, it is not noticed that the value that poses the self referent case belongs to such a subclass, it may happen that a pattern of inference which is being assumed to hold for all the values of the variable, does not in fact hold for the value that poses the self referent case. We thus get the fallacy that the variable is conceived too narrowly to permit the self referent case at the same time that this case is permitted. The problem, then, is whether or not the very conception of the function is compatible with enlarging the scope of the variable to permit the self referent case. But of course such a reinterpretation of the function becomes urgent only if its self referent use must be regarded as justified.

Russell's Antinomy provides a clear example of the fallacy in question and of the possibility of reinterpreting the function to permit the self referent case. But as for the 's_v' of the Lukasiewicz Paradox, we find that the very conception of this propositional

* It is noted that this hypothesis might perhaps be developed in a more fundamental way in the terms of a logical group theory in which the group is associated with a universe of discourse; but in view of its great difficulties, this focus on the problem is not adopted here. —See Note 15 of the paper.

form seems to be incompatible with its valid self referent use. In what remains of this summary, I shall merely indicate how this matter seems to me to work out in the respective problems.

With respect to the α and β of Russell's problem, the pairs of functions $(\hat{W}x, \sim\hat{W}x)$ and $(\hat{X}x, \sim\hat{X}x)$ are, of course, contradictories; yet this is compatible with there being a subclass of classes for which the reflexive class membership relation does not hold. Contradictory functions are such on the assumption that their use is within the domain of true or false discourse. Then if the value w of x takes the use of these functions out of this domain, with the result that $\hat{W}w$ and $\sim\hat{W}w$ are neither true nor false (as I have suggested in *Part II*), this does not refute that $(\hat{W}w, \sim\hat{W}x)$ and $(\hat{X}x, \sim\hat{X}x)$ are contradictory pairs when used within the domain of true or false discourse. However, to develop this solution requires that we revise our conception not only of the class variable, to include the subclass of classes to which w belongs, but also of the functions α and β. Assuming (as suggested at the end of *Part II*) that the strange objects which have cropped up are imaginary, we could say that these functions contain constant imaginary components —most notably, "W" itself— and thus admit only of complex or wholly imaginary use. On this view, when x is a real class, their use is complex and within the domain of true or false discourse; but when x is imaginary, their use leaves this domain. If this way of handling the problem were adopted, then, among the remaining difficulties, there would be a need for a special symbol (or perhaps more than one) to designate the imaginary components of expressions.

Meanwhile, it should not be overlooked that to regard $\hat{W}$ as an imaginary class membership condition involves exploiting the systematic ambiguity of the membership condition property variable $\hat{X}$. We found in *Part II* that "W" is the mere general idea of the membership condition of w. Now we see that, regarded as a value of $\hat{X}$, $\hat{W}$ is a very strange kind of property. Just as there is no square root of —1, there is no univocally specified property $\hat{W}$. There is nothing but the idea of it; and in order that $\hat{X}$ may take $\hat{W}$ as a value, the conception of the logical property variable has to be pushed, so to speak, to a limit of sense.

On the ground of these results it is concluded that Russell's Antinomy is resolvable on the basis of accepting his preliminary argument, if it is conceded that he discovered a new class of logical objects. Moreover, this type of solution, if it is tenable, is surely desirable. The great strength of this paradox consists precisely in that its impeccable and intuitively valid preliminary argument, lead-

ing to the conception of the function, discovers a problem whose solution clearly requires a development of logical theory Thus no solution that recurs to a theoretical development which in its turn invalidates the preliminary argument, can wholly dispel the logical disquiet to which this paradox has given rise.

9. We now turn our attention to the workings of self reference in the case of the singular self referent sentence. We note first that by any method of constructing such a statement, *the sentence that the statement is about is actually present, in the role of an indicated object,* and in this role has the essential ambiguity of any object that can be pointed at. It being a linguistic object, a sentence, what we can make of this object is first of all a matter of what kind of sentence we can take it to be. For example, the moment we focus on the self referent sentence as a statement about a sentence, we see that if its predicate is applicable to a verbal sentence, then such a self referent sentence is interpretable as being about the verbal sentence used to state it. Moreover, this interpretation does not involve regression, because, logically, the verbal sentence is a propositional form whose subject term is a variable and as such has no referent. (See Note 16.) Another possibility is that if its predicate is applicable both to statements and verbal sentences (e.g., "well formed", "significant", etc.), a self referent sentence is interpretable as being about a statement which is in turn about a verbal sentence. Still another possibility is that a self referent sentence may be stipulated to be a non-regressive statement about a sentence that involves an infinite regression. This interpretation is grounded in the consideration that the main statement made by the self referent sentence, in so far as it is a statement, could in principle be asserted in different words. Examples of self referent sentences that can be interpreted significantly in these ways are given in the body of the paper.

The riddle posed by the self referent sentential construction is simply to determine in how many different ways this kind of construction can exploit the systematic ambiguity of the term "sentence". Depending upon its predicate, the self referent sentence can be used to express something significant about one or another kind of sentence. So we have here again the exploitation by self reference of the systematic ambiguity of a variable term.

It remains to consider whether the 's_V' of the Lukasiewicz Paradox, which has the form of a truth locution, admits of self referent use.

Now it has been shown that the circular 's' of the paradox is not a well formed sentence. This use of 's_V' commits the fallacy of

permitting a value of its subject term by which the subject-predicate form of 's_v' is not conserved; and the result is that *no* pattern of inference for reasoning from a sentence in the subject-predicate form is applicable to this 's'.

Next. No matter how one tries to reinterpret the statement 's', it seems to turn out at best to be a category mistake. Its predicate being one of the complementary pair, ("true", "not-true"), it is impossible to enlarge the scope of the variable subject term of 's_v' to permit as values those classes of sentences which are neither true nor false, whether they be verbal sentences, or category mistakes, or infinite regressions, or vicious circles, or whatever. Nor does there seem to be any way of interpreting 's' to be about a true or false statement. Thus is may safely be concluded that, as far as one can see, there is no significant, self referent use of 's_v'; and it follows that there is no such thing as a singular, self referent truth locution.

Does this, then, imply that the use of the predicate "true" fails within natural language? Indeed not. To be sure, if there were a reason that compelled us to use 's_v' self referently, then there would be a conflict of criteria and a paradox. But in the absence of such a reason, we may simply conclude that the self referent use of 's_v' involves a use of "true" that is a category mistake. Nor is this at all odd, since it is notorious that the use of "true" is peculiarly exposed to this fallacy.

For the rest, it is surely evident that in so far as its expressive value is concerned, the singular self referent sentence is a mere curiousity. Nonetheless, this logicians brainchild has a different kind of utility. By confronting a statement with the verbal sentence used to make it, this construction provides a highly suggestive context for the further study of the distinction between the statement and he verbal sentence. Thus we see once again how self reference can be useful to logic itself.