

COMPRENDER AL OTRO ES CREAR UN ESPACIO COMPARTIDO

Caridad, empatía y triangulación

PABLO QUINTANILLA PÉREZ-WICHT
DEPARTAMENTO DE HUMANIDADES
PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
pquinta@pucp.edu.pe

Resumen:

En este artículo se analiza la versión clásica del principio de caridad propuesta por Davidson, para sugerir una reformulación de la que se desprenden algunas consecuencias sobre la naturaleza de la comprensión. Así, se muestra un alejamiento de la hermenéutica intencionalista y de la concepción cartesiano-internalista de la mente, para sugerir que la comprensión es el proceso creativo en el que los interlocutores generan un espacio compartido de disposiciones comportamentales, estados mentales, y valoraciones. La generación de este espacio compartido es propiamente la construcción de una forma de vida, la cual es el trasfondo, no siempre consciente, desde el cual justificamos nuestras convicciones epistemológicas y morales. Esto exige integrar el principio de caridad con la noción de empatía o simulación, como el criterio requerido para la atribución de estados mentales, así como para explicar la divergencia entre ellos.

Palabras claves: Davidson, principio de caridad, racionalidad, comprensión, empatía.

Abstract: *Understanding the Other is Creating a Shared Area*

This paper analyzes Davidson's classical version of the principle of charity, in order to suggest a reformulation from which it is possible to extract some consequences regarding the nature of understanding. Thus, it is put forward an abandonment of the intentionalist hermeneutics as well as the Cartesian-internalist conception of the mind, in order to suggest that understanding is the creative process in which both interlocutors generate a shared area of dispositions to act, mental states, and values. The generation of this shared area is precisely the construction of a form of life, as the background from which we justify our epistemological as well as moral convictions. This requires to integrate the principle of charity with the notion of empathy or simulation, as the criterion for the attribution of mental states, as well as for the explanation of the divergence among them.

Key words: Davidson, principle of charity, rationality, understanding, empathy.

El principio de caridad constituye uno de los desarrollos del pensamiento de Davidson que más interés han concitado en la filosofía reciente. No sólo es relevante para temas específicos en filosofía del lenguaje, epistemología y filosofía de la mente, sino también tiene im-

portantes consecuencias para la reflexión teórica en disciplinas hermenéuticas, como la psicología y las ciencias sociales. Para percibir esos alcances, sin embargo, es necesario ver al principio de caridad como un conjunto de tesis que va mucho más allá de la simple metodología de la interpretación.

En este texto me propongo sugerir una formulación del principio de caridad a partir de la cual es posible extraer consecuencias acerca de la naturaleza de la comprensión. Deseo mostrar que esta formulación hace explícito el abandono de la hermenéutica intencionalista así como de la concepción internalista de la mente, aquellas según las cuales comprender al otro es conocer sus intenciones y otros estados mentales internos, y más bien permite entender la comprensión como la creación de un espacio compartido entre intérprete y agente. Así, entender al otro no sería reconstruir su vida mental, ni conocer los estados mentales que causaron sus acciones, sino generar una forma de vida común, una intersección de horizontes que conlleva a una invitación al cambio, la tolerancia y la adaptación.¹

Hay muchas formulaciones diferentes del principio de caridad, pero la formulación clásica de Davidson aparece en el siguiente párrafo:

En nuestro afán de hacer que tenga sentido lo que dice [el hablante], intentaremos con una teoría que lo encuentre consistente, un creyente de verdades y un amante del bien (todo de acuerdo con nuestros propios estándares, sobra decirlo). (Davidson 1980: 222)²

El problema con las formulaciones davidsonianas del principio de caridad es que tienden a ser imprecisas y no explicitan suficientemente sus detalles y presupuestos. Esto ha originado cuestionamientos de fondo. En algunos lugares Davidson dice que el objetivo del principio de caridad es optimizar el acuerdo (*cf.* Davidson 1984: 197) o minimizar el desacuerdo (*Id.*, 1984: xvii). Más adelante aclara esas formulaciones y sostiene que:

¹ Las alusiones a Wittgenstein y a Gadamer muestran que hay muchas conexiones que es posible hacer entre estos autores y Davidson. Algunas de esas asociaciones las he desarrollado en Quintanilla 1999 y 1997.

² Se pueden encontrar otras formulaciones de este principio en Davidson 1980: 221, 290; así como en Davidson 1984a: 137; Davidson 1984b: 153; Davidson 1984c: 168s; Davidson 1984d: 196s. Todas ellas tienden a ser una elaboración del *dictum* quineano: "la estupidez de nuestro interlocutor, más allá de cierto punto, es menos probable que una mala traducción" (Quine 1960). El principio de caridad fue propuesto por primera vez por N.L. Wilson (Wilson 1959: 432; Wilson 1970: 300) como parte de una teoría de la referencia. Posteriormente fue formulado de una manera más amplia por Quine en su análisis de la traducción radical en Quine 1960: 57s.

La meta de la Interpretación no es el acuerdo sino la comprensión. Mi punto siempre ha sido que la comprensión puede ser asegurada solo interpretando en una forma que sirva para el tipo de acuerdo adecuado. El “tipo adecuado”, sin embargo, no es más fácil de especificar que decir lo que constituye una buena razón para sostener una creencia particular. (Davidson 1984: xvii)

En este texto intentaré proponer una formulación más precisa del principio de caridad que sea, al mismo tiempo, más hábil en defenderse de los diversos cuestionamientos sugeridos. Para ello, haré uso de la vieja noción de empatía, que en la filosofía reciente ha sido reformulada como capacidad de ‘simulación’. También trabajaré sobre la noción de triangulación, que Davidson y Marcia Cavell utilizan, aunque es una noción psicoanalítica en su inspiración (cf. Davidson 2001 y Cavell 1993, 1998).

La concepción de empatía, desarrollada por Schleiermacher y Dilthey, y que se puede rastrear hasta el concepto de *sympathy* en Hume,³ alude a la capacidad de ponerse en el lugar del otro para padecer con él sus venturas o desdichas. Este concepto fue duramente cuestionado por estar aparentemente comprometido con una hermenéutica intencionalista. También ha sido criticado por estar comprometido con un modelo *transportacional* de la comunicación,⁴ según el cual comprender al otro es ser capaz de conocer sus estados mentales, su escenario interior para ponerlo en términos de Gilbert Ryle, mediante la *transportación* de esos estados desde su mente a la nuestra. Aunque Locke es el defensor paradigmático del modelo transportacional (cf. Locke 1975: lib. 3 cap. ii §1), también podemos encontrarlo en la teoría de la empatía de Dilthey: “De la riqueza total de nuestra propia experiencia, la experiencia exterior es reproducida en nosotros y comprendida por transportación” (Dilthey 1913-1917: vol. 5, 263). Parece pues que las formulaciones clásicas de la empatía están comprometidas con una concepción internalista-cartesiana de la mente, donde comprender al otro es reproducir en nuestro escenario interior, por transportación, lo que creemos que está ocurriendo en su propio escenario interior.

³ Hume pensaba que en el origen de la moral se encuentra la empatía, en tanto nos permite alejarnos de nuestra perspectiva e interés para comprender la perspectiva e interés de otra persona. Sostenía que la empatía está presente en la comprensión de los demás tanto como en los juicios estéticos y éticos. Pensaba que tenemos también una tendencia natural a *proyectar* nuestros sentimientos morales y estéticos sobre las personas y las cosas, “spreading our minds upon nature”. Cf. Hume 1975.

⁴ Algunos, como Jonathan Bennett (Bennett 1971), denominan a este modelo como de la *traducción*, pero yo prefiero denominarlo transportacional para no confundirlo con los modelos de traducción radical de Quine. La expresión ‘modelo transportacional’ ha sido acuñada por Robert Brandom en Brandom 1994.

Quizá por esos motivos, en la filosofía de la mente reciente así como en la psicología experimental se ha acuñado el concepto de ‘simulación’, que se entiende como la capacidad para imaginar ser el otro en condiciones contrafácticas. La teoría de la simulación, que pretender explicar la manera como el niño desarrolla una ‘teoría de la mente’ y así puede imaginar ser el otro mediante el uso de la capacidad de la meta-representación, ha sido extensamente estudiada en la bibliografía actual (cf. Davies/Stone 1995a y 1995b; Whiten 1991; Astigton et. al. 1988). Naturalmente, otra opción es mantener el concepto de empatía en un sentido no-intencionalista. No voy a detenerme ahora en las diferencias que pudiera haber entre empatía y simulación, porque para los propósitos de este artículo no son relevantes, sólo subrayaré que emplearé una noción de empatía o simulación no intencionalista. Tampoco entraré en este artículo en el debate sobre si uno simula *ser* el otro bajo condiciones contrafácticas o simula *tener* los estados mentales del otro bajo condiciones contrafácticas, aunque preferiré la primera opción.

Sostendré que el objetivo del principio de caridad es minimizar la divergencia inexplicable e intentaré mostrar que el criterio para la explicación de la divergencia debería ser la noción de empatía. La idea será que atribuimos al agente nuestros estados mentales,⁵ a menos que tengamos razones para creer que él no puede tener esos estados mentales. Así, debemos tener un criterio para determinar cuáles son sus estados mentales idiosincrásicos y por qué él los tiene mientras que nosotros no; o por qué nosotros tenemos estados mentales que él no tiene. Como he señalado, nuestro criterio será la empatía. Imaginamos cómo es ser él en una situación en particular y atribuimos al agente los estados mentales que creemos que nosotros tendríamos si fuéramos él en esa circunstancia. En lo que sigue expondré las nociones de triangulación y empatía, y mostraré cómo se pueden integrar con una reformulación del principio de caridad.

La triangulación es un fenómeno que ocurre cuando dos interlocutores reaccionan al mismo tiempo a lo que ellos consideran que es su mundo compartido, reaccionando cada uno a las reacciones del otro. En el caso de la interpretación radical, la intérprete⁶ es consciente de que ciertas circunstancias de su entorno causan que el agente produzca cierta proferencia verbal *p*. Sin embargo, la intérprete no sabe cuál es el significado de *p* para el agente ni cuáles son las creencias causa-

⁵ En lo que sigue, consideraré “estados mentales” a las creencias, los deseos e incluso a las acciones, en tanto surgen de una situación interpretativa intersubjetiva. Asumiré que los estados mentales no tienen un origen ni una realidad subjetiva sino que, por el contrario, se constituyen intersubjetivamente en la situación interpretativa.

⁶ Por razones de comodidad estilística, me referiré a la intérprete en femenino y al agente en masculino.

das en él por esas circunstancias. Pero la intérprete sí sabe qué creencias han sido causadas en ella en las mismas condiciones. Así, la intérprete atribuirá al agente las creencias que esas condiciones han causado en ella. En otras palabras, ella atribuirá al agente algunas de sus creencias y asignará como el significado de *p* en el lenguaje del agente una oración de su propio lenguaje con las mismas condiciones de verdad.⁷

Davidson rechaza la idea de que la intérprete *proyete* sus estados mentales en el otro, por considerar que esto tiene una connotación anti-realista (cf. Davidson 1997). Discrepo con él en lo relativo a las connotaciones anti-realistas. La intérprete puede proyectar sus propios estados mentales precisamente con la finalidad de encontrar los del otro, los cuales tienen realidad independiente de la intérprete aunque son constituidos en circunstancias intersubjetivas, pero la evidencia siempre subdeterminará los distintos manuales de interpretación que se puede hacer de ellos. En otras palabras, los estados mentales del otro son reales pero indeterminados. Sin embargo, es verdad que el uso del término 'proyección' puede sugerir una concepción demasiado egocéntrica de la interpretación. Volveré sobre este punto más adelante.

En general, la intérprete asumirá que el agente reaccionará frente a los estímulos externos básicamente de la misma manera como ella reaccionaría. Considérese a ambos interlocutores reaccionando frente a los mismos estímulos, una manzana, por ejemplo. La manzana causará percepciones en ambos interlocutores y cada uno de ellos observará las reacciones del otro. Así, cada uno correlacionará sus propias percepciones y creencias, en relación a la manzana, con los del otro. Lo importante es que cada interlocutor reaccionará frente al mundo y también frente a las reacciones que el otro tiene respecto del mundo. Cada interlocutor deberá ver las reacciones ajenas como una perspectiva *diferente* del *mismo mundo*. De esta manera, cada uno asociará el concepto de manzana con sus propias reacciones y con las reacciones del otro. Al hacerlo, cada interlocutor aprenderá el concepto 'manzana' de manera intersubjetiva, asumiendo que hay un uso objetivo que este concepto tiene, y atribuirá al otro creencias y deseos relacionados con el objeto y con el concepto 'manzana'. Más aún, cada uno de ellos aprenderá a asociar las reacciones del otro con sus propias reacciones, y así también aprenderá a simular ser el otro bajo circunstancias similares; es decir, aprenderá a empatizar con el otro en relación a un mundo que asumen compartido.

Ahora bien, la intérprete podrá comenzar a interpretar al agente sólo si ella tiene dos conjuntos de supuestos iniciales acerca de él. De

⁷ Esto se formula de manera técnica como la producción de oraciones-T, en que: (T) "*s*" es verdadera en *L* si y sólo si *p*, donde *s* está en el lenguaje objeto y *p* en el metalenguaje, teniendo ambas las mismas condiciones de verdad.

un lado, un conjunto de creencias de primer grado en torno del agente y del mundo que comparten; de otro lado, un conjunto de creencias de segundo grado acerca de las creencias del agente, así como acerca de sus deseos y propósitos al realizar sus acciones. La intérprete formará estas hipótesis interpretativas al considerar el comportamiento del agente tanto como su entorno, y se formará creencias acerca de los estados mentales del agente al atribuirle parte de su propio sistema de estados mentales. Pero en este punto, la intérprete necesitará un criterio para determinar cuáles estados mentales deberán ser atribuidos a él. Ese criterio es la empatía. Al simular ser él, ella intentará ver el mundo desde la perspectiva que ella cree que él tiene. Así, si la interpretación comienza cuando ella le atribuye a él un aspecto de *su propio* sistema de estados mentales, ella tendrá que asumir que, en líneas generales, él cree lo que es verdadero y desea lo que es deseable, todo según los criterios de la intérprete. Este, por supuesto, es el principio de caridad.

Habrà que recordar que el principio de caridad no está haciendo una sugerencia metodológica para la interpretación; el principio de caridad sostiene que es condición de posibilidad de la interpretación que la intérprete asuma que *en líneas generales*:

- (i) El comportamiento del agente es consistente.
- (ii) El agente mantiene los mismos supuestos lógicos que la intérprete.
- (iii) Las creencias del agente son verdaderas.
- (iv) El agente se comporta siguiendo su mejor juicio. Esto es, el agente persigue los mismos fines que la intérprete al realizar sus acciones bajo las mismas condiciones generales.

La justificación principal del principio de caridad toma la forma de un argumento trascendental, en tanto muestra que ciertas prácticas de tipo A (la atribución de estados mentales a agentes intencionales) presupone *necesariamente* ciertas otras prácticas de tipo B (la aplicación de las condiciones del principio de caridad). Ya que no podemos evitar estar comprometidos con A, entonces también debemos estar comprometidos con B.⁸

Al justificarse en un argumento trascendental, el principio de caridad se muestra como una condición necesaria para la interpretación y para la atribución de estados mentales, no es una afirmación empírica acerca del comportamiento de los intérpretes. Si fuese una afirmación empírica, podría falsearse mediante contraejemplos empíricos. Así pues, ningún descubrimiento antropológico podría mostrar que es falso, sólo un análisis más fino de los conceptos de interpretación. El

⁸ Hay una copiosa bibliografía acerca de la validez de los argumentos trascendentales. No voy a entrar en el debate ahora, aunque en líneas generales asumiré su validez si el argumento está bien formulado. Para el debate, cf. por ejemplo Hintikka 1972, Bieri et. al. 1979 y Malpas 1990.

punto es que de no asumir el principio de caridad, no podríamos reconocer comportamiento intencional en las personas y, por tanto, no podríamos reconocerlos como *agentes* y no podríamos atribuirles creencias y deseos en absoluto.⁹

Analizaré ahora cada una de las condiciones del principio de caridad señaladas líneas arriba. *La condición (i)* sostiene que debemos asumir consistencia general en el comportamiento del agente si esperamos reconocerlo como comportamiento intencional. La consistencia debe ser entendida tanto en un sentido lógico como en un sentido pragmático. En un sentido lógico, la atribución de demasiadas contradicciones cuestiona la posibilidad de adscribir un sistema de estados mentales interconectados, así como la posibilidad misma de dar contenido a cada uno de ellos. La intérprete sólo podrá determinar el contenido de cada estado mental al interconectarlo con los otros estados mentales atribuidos al sistema.¹⁰ Pero en un sentido pragmático, el comportamiento de un agente es consistente si sus estados mentales y sus acciones no son incompatibles según la perspectiva de la intérprete. Serían incompatibles si se excluyeran mutuamente. Naturalmente, hay casos de irracionalidad en que al agente se le atribuyen estados mentales inconsistentes. Sin embargo, incluso en estas situaciones, la atribución de irracionalidad se hace con el fin de seguir encontrándolo básicamente consistente, es decir, se hace para que pueda seguir siendo inteligible a los ojos de la intérprete. Pero no entraré ahora en la naturaleza de la irracionalidad, que es un tema en sí mismo y que requiere de la atribución de estados mentales inconscientes.

Dado que el concepto de creencia es fundamental en esta exposición, será necesario delimitarlo. En principio, es posible definir la creencia de muchas maneras, pero el concepto de creencia que ahora nos interesa es el que la intérprete deberá asumir para atribuir creencias al agente. Sospecho que necesitamos una concepción de la creencia inspirada en Charles Sanders Peirce, según la cual creer que *p* es tener la disposición para comportarse como si *p* fuese verdadera. A ello habrá que añadir la disposición para aceptar aquellas otras creencias que se deriven de *p*, y, finalmente, lo que Davidson llama *el concepto de creencia*, esto es, la capacidad para reconocer la diferencia entre una creencia verdadera y una falsa. Esta concepción de creencia

⁹ Cf. David Lewis: “[una persona] podría no tener creencias, deseos o significados en absoluto pero es analítico que si las tiene entonces éstas más o menos se conforman a los principios regulativos por los cuales los conceptos de creencia, deseo y significado son definidos”. (Lewis 1983: 12, cf. Lewis 1972)

¹⁰ Así también es la manera como aprendemos nuevas creencias; no lo hacemos una por una sino mediante pequeños subsistemas. Cf. Wittgenstein: “cuando comenzamos a creer algo, lo que creemos no es una proposición singular, es todo un sistema de proposiciones (la luz alborea gradualmente sobre la totalidad)”. (Wittgenstein 1972: 41)

tiene la virtud de mostrar que el razonamiento teórico y el práctico es el mismo. Si adoptamos una definición de creencia como la descrita, sería imposible que alguien se comportara, de manera sistemática, de forma contraria a sus creencias, precisamente porque las creencias son construcciones culturales que la intérprete atribuye al agente para poder hacerlo inteligible.

La condición (ii) sostiene que, ya que la intérprete atribuye al agente parte de su propio sistema de estados mentales, no podrá evitar reconocer el comportamiento del agente desde su propia estructura lógica.

La condición (iii) sostiene que la intérprete debe asumir que las creencias del agente son básicamente verdaderas, es decir, que son semejantes a las de ella. Esto se sigue de la necesidad de atribuir, mediante el proceso de la triangulación, un sistema integrado de creencias, donde las interacciones entre éstas fijan el contenido de cada una de ellas. En este sistema, la falsedad debe ser excepcional.¹¹ Un creyente sistemático de falsedades no podría ser triangulado con el entorno. Tampoco podría determinarse cuáles son sus creencias falsas, porque la intérprete sólo podría precisar la falsedad de una creencia del agente contra el telón de fondo de creencias verdaderas que le permiten reconocer su significado. Así pues, una intérprete no podría creer que su interlocutor es un creyente sistemático de falsedades a costa de que se evaporen los conceptos mismos de 'creencia' y 'falsedad'. Pero no sólo eso, mediante el uso de un argumento trascendental semejante al que justifica el principio de caridad, podría afirmarse de manera más general que sería imposible que alguien fuese un creyente sistemático de falsedades. El argumento tomaría más o menos esta forma: puesto que todo sistema de creencias se construye en una situación interpretativa en la que la intérprete lo atribuye al agente para hacerlo inteligible, y como sería imposible que alguna intérprete construya para su interlocutor un sistema constituido principalmente por creencias falsas, entonces sería imposible que existiese un sistema constituido principalmente por creencias falsas. No podría ocurrir que alguien tenga un sistema de creencias principalmente falsas independientemente de toda interpretación, porque al afirmar esa posibilidad se estaría negando la tesis central del principio de caridad de que los estados mentales se atribuyen de manera intersubjetiva y no son propiedades monádicas de mentes individuales. Así pues, el principio de caridad tiene como consecuencia afirmar una propiedad de los sistemas de creencias como tales: pertenece al concepto mismo de creencia que la mayor parte de creencias de un sistema deben ser verdaderas.

Pero ahora el problema es llegar a formular el principio de caridad de una manera rigurosa y precisa. Podríamos hablar de maximización del acuerdo y racionalidad, y minimización de falsedad e irracionalidad. Así, "cuando la formulación habitual que hace Davidson (y Davidson sus errores. Demasiadas equivocaciones simplemente distorsionan el foco" (Davidson 1984c: 168).

son 1986: 309). Pero en este caso habría que aclarar qué clase de acuerdo debe asumir la intérprete, esto es, acuerdo acerca de qué tipo de creencias; además, habría que establecer el criterio que ella tendría que emplear para explicar las divergencias de creencias entre ella y el agente. Davidson no ofrece respuestas a esta pregunta y tampoco tiene un criterio para la explicación de la divergencia. Por ello me parece preferible decir que el objetivo del principio de caridad es minimizar las divergencias inexplicables de estados mentales entre la intérprete y el agente, usando como criterio la noción de empatía.

Hemos visto que la intérprete comenzará la interpretación atribuyendo al agente sus propios estados mentales, pero ella necesitará un criterio para determinar cuáles de sus estados mentales deberá atribuirle. Ella le atribuirá aquellos estados mentales que ella cree que tendría si fuera el agente, en las circunstancias en que él se encuentra. Al hacer eso, intentará ver el mundo desde el punto de vista del otro. De esta manera, no sólo le atribuirá muchas de sus propias creencias sino también creencias que ella no tiene o que considera que son falsas. Las creencias sobre las cuales la intérprete debe asumir acuerdo, tendrán que ser aquellas que la intérprete considere que debe atribuir al agente para poder hacerlo inteligible. Y, por supuesto, no hay manera de saber por adelantado cuáles serán esas creencias.

Al asumir que las creencias del agente son básicamente verdaderas, la intérprete supone también que él está básicamente acertado en la manera en que el mundo es y, por tanto, que ellos comparten un mundo que es previo e independiente de ellos. Aquí también tenemos un argumento trascendental que prueba que sería imposible *no* creer en la objetividad e independencia de la realidad. El argumento no prueba que exista un mundo previo e independiente de nosotros, lo que prueba es que es condición necesaria para la interpretación que asumamos eso. Ahora bien, ya que sería imposible creer lo contrario, hemos probado que es necesario que creamos que el mundo es previo e independiente de nosotros. Esto es todo lo que se necesita probar en relación a la existencia de la realidad. Así satisfacemos nuestras intuiciones realistas, que no podríamos dejar de tener, y también dejamos satisfecho a quien considera que conocemos al mundo desde una cierta interpretación que ya tenemos de él, de suerte que es imposible siquiera concebir el mundo si no es interpretado de alguna forma. Esto resulta importante porque justifica una tesis realista mediante un argumento trascendental basado en la naturaleza misma de la interpretación y no en un argumento metafísico. Esta posición podría ser denominada, como lo hace Hubert Dreyfus en otro contexto (Dreyfus 1996), realismo hermenéutico. Al triangular entre sus propios estados mentales, los del otro, y los objetos del mundo que consideran compartir, los interlocutores asumirán que su acuerdo procede de la objetividad misma del mundo que habitan.

Richard Grandy ha ofrecido una versión diferente del principio de caridad, a la cual ha llamado “principio de humanidad” (Grandy 1973).¹² Este principio no tiene como objetivo maximizar el acuerdo sino minimizar el desacuerdo inteligible o, lo que es lo mismo, maximizar la inteligibilidad aunque no necesariamente el acuerdo. Hay quienes creen que el principio de humanidad es un avance respecto del de caridad al implicar que puede haber inteligibilidad sin acuerdo (*cf.* Papineau 1978). Graham Macdonald también cree que deberíamos sustituir humanidad por caridad, dice:

Más que tomar la caridad al extremo de minimizar todo desacuerdo, una estrategia interpretativa más razonable parecería ser ocuparse solamente de minimizar el desacuerdo inteligible. (Macdonald 1981: 37)

A primera vista estas objeciones son sorprendentes, porque parecen implicar que puede haber inteligibilidad sin un fondo de acuerdo. Precisamente, el punto del principio de caridad es intentar probar que sólo puede haber inteligibilidad y desacuerdo sobre un fondo de acuerdo. Pero lo que Grandy en realidad está sugiriendo es que, aunque debe haber un fondo de acuerdo, algunas veces podemos hacer más inteligible al agente si le atribuimos un subsistema de creencias falsas, siempre que podamos explicar cómo es que él las adquirió. Así pues, Grandy objetaría una formulación del principio de caridad que sostuviera, como lo hace Quine, que la mejor traducción es aquella que maximiza el acuerdo. El problema con la formulación de Quine es que no explica cómo en ocasiones la atribución de falsas creencias puede ayudar a hacer más inteligible al agente. La posición de Quine intentará minimizar todo lo posible la atribución de creencias falsas, mientras que Grandy sugeriría minimizarlas sólo cuando no podamos explicar cómo fueron adquiridas estas creencias.

Al defender la maximización del acuerdo en la interpretación, [el principio de caridad] aconseja el abandono de consideraciones acerca de si los interpretes tienen la posibilidad de haber alcanzado conocimiento de las verdades sobre las cuales ellos son interpretados como estando de acuerdo. (Grandy 1973: 29)

Coincido con Grandy en este punto. El principio de caridad debería apuntar a minimizar el error y la irracionalidad, sólo cuando sean inexplicables. La pregunta es cuál será nuestro criterio para la atribución de creencias específicas al agente. Grandy sugiere tomarnos a nosotros mismos como modelos, pero no llega a desarrollar una noción de empatía o simulación.

¹² Aquí Grandy comenta la formulación de Quine del principio de caridad, ya que la formulación de Davidson no había aparecido aún.

Entonces tenemos, como una constrictión pragmática sobre la traducción, la condición de que el patrón atribuido de relaciones entre creencias, deseos y el mundo sea tan similar al nuestro como sea posible. [...] El que nuestra simulación de la otra persona sea exitosa dependerá fuertemente de la similitud entre su red de creencias y deseos y la nuestra. (*Id.*, 43)

Ahora, hacer al agente lo más semejante a uno podrá ser proyección, pero no puede ser simulación o empatía, precisamente porque la idea de la empatía es el intentar alejarnos de nuestros propios estados mentales para imaginar ser el agente, dejando espacio en nosotros mismos para albergar al otro en su particularidad. Así pues, la versión que necesitamos del principio de caridad es una que permita a la intérprete la flexibilidad suficiente para desplazarse más allá de sus propios estados mentales imaginando los del otro. Ella no sólo deberá atribuir una parte de sus propios estados mentales al agente sino que también deberá imaginar nuevos, incluso inusuales o exóticos. Pero para poder hacer eso, ella deberá ser capaz de atribuirle a él aquellas creencias que ella piensa que estaría en condiciones de adquirir si fuera el agente. Posteriormente y dependiendo del comportamiento del agente, ella tendrá que hacer los ajustes necesarios para atribuir nuevos estados mentales. Intentaré aclarar este punto al contestar una objeción de Michael Devitt.

Devitt nos invita a imaginar a un intérprete no religioso tratando de comprender a un nativo involucrado en algunos ritos religiosos. La aplicación del principio de caridad, piensa Devitt, no permitiría hacer inteligible al nativo. Dice Devitt:

¿Qué voy a hacer con las con las partes religiosas de su conducta no verbal? Éstas parecerán inexplicables. Interpreteselo no caritativamente, sin embargo, y esa conducta caerá en su lugar. (Devitt 1981: 116)

Si solo proyectamos en el agente lo que consideramos verdadero utilizándonos a nosotros mismos como modelo, jamás comprendemos adecuadamente al agente. En efecto, ese modelo es demasiado egocentrista para poder ser correcto: convierte al agente en un desdoblamiento de la intérprete. Precisamente por eso es que la simulación se hace necesaria.

Veamos ahora cómo ocurriría toda la situación interpretativa. En primer lugar, la intérprete tiene que atribuir al agente aquellas creencias que ella considera obviamente verdaderas acerca de su mundo compartido. Luego de observar el comportamiento del agente y las nuevas circunstancias, la intérprete deberá atribuirle aquella creencia que ella piensa que tendría si fuera él. De esta manera, si ella observa al agente realizando un tipo de comportamiento que parece presupo-

ner creencias falsas (imaginemos, por ejemplo, que él está adorando al sol), ella deberá imaginar cómo sería tener creencias verdaderas acerca de las características evidentes del entorno físico pero no creencias especializadas acerca de las características del sol, mientras también se tiene creencias verdaderas acerca de la importancia del sol para la vida humana. De aquí podemos encontrar razonable y hasta casi predecible que el agente crea que el sol es una suerte de divinidad que debe ser adorada. La intérprete atribuirá al agente algunas creencias falsas acerca del sol, lo que le permitirá encontrar la racionalidad de su comportamiento, permitiéndole a ella comprenderlo mejor.

No sólo atribuimos a los demás nuestras creencias sino también nuestro concepto de creencia, y nuestra convicción de que el concepto de creencia involucra necesariamente la suposición de que las creencias pueden ser verdaderas o falsas, al punto de que creer que *p* entraña creer que *p podría* ser falsa. Pertenece al concepto de creencia que uno asuma que sus creencias son verdaderas, pero también pertenece al concepto de creencia el creer que esas mismas creencias *podrían* ser falsas. Asumir esto es poseer los conceptos de subjetivo y objetivo: esto es, que hay una manera como las cosas son que es independiente de nuestra voluntad. Por ello, la intérprete deberá atribuir al agente al mismo tiempo un sistema de creencias verdaderas y el concepto de creencia, así como la capacidad para distinguir entre lo subjetivo y lo objetivo, es decir, una ontología realista y la creencia en que hay una verdad objetiva que no depende de las voluntades individuales.

Pero recuérdese que no se está sosteniendo que *existe* una verdad objetiva que es independiente de las voluntades individuales; lo que se sostiene es que es necesario que lo presupongamos para que la interpretación, y todo lo que viene con ella, sea posible. Sin embargo, esto puede funcionar como un argumento trascendental, porque si es necesario para la interpretación que asumamos que existe una verdad objetiva, y como no podríamos evitar ser intérpretes, entonces no podemos evitar creer que existe una verdad objetiva. Por otra parte, no puedo evitar comportarme como si existiese una verdad objetiva. Luego, si creo que existe una verdad objetiva, estoy lógicamente obligado a sostener que existe una verdad objetiva. Estas conclusiones objetivistas y realistas, mediante un argumento trascendental, son conclusiones inevitables del principio de caridad. Pero todo esto es posible sólo si la intérprete puede triangular entre ella, el agente y el mundo que ambos comparten y que consideran independiente y previo a ellos.

Ahora bien, como he intentado sostener, una intérprete podrá atribuir a un agente una creencia *p* sólo al interior de un sistema mucho mayor de creencias y otros estados mentales interconectados con *p*. Ya que la definición de creencia que estoy empleando incluye, aunque no se reduce a, una disposición para actuar, entonces un sistema de creencias compartido por intérprete y agente puede ser visto como un

sistema compartido de disposiciones para comportarse, o un sistema de prácticas sociales que constituyen una forma de vida. Con frecuencia estas disposiciones no son plenamente conscientes para los agentes mismos, porque existen a un nivel pre-conceptual de interacción con el mundo y con las otras personas. Sin embargo, lo que deseo sugerir es que la atribución de una creencia cualquiera implica la atribución de una forma de vida asumida como compartida. Podría objetarse que atribuimos creencias a bebés o niños muy pequeños incluso si no los consideramos todavía miembros de nuestra forma de vida. Sospecho que si pensamos que podemos atribuir creencias así como el concepto de creencia a un niño pequeño, deberemos también poder verlo como miembro de nuestra forma de vida, incluso si es en un sentido muy rudimentario. Sin embargo, es muy probable que la transición desde, de un lado, habilidades tácitas y pre-conceptuales hacia, del otro, creencias y prácticas sociales, se realice de manera gradual y progresiva, donde no hay una frontera entre estos dos extremos sino sólo un continuo.

La condición (iv). Las creencias que la intérprete asume que comparte con el agente son creencias acerca de los objetos comunes, acerca de los significados de las palabras, y acerca de los que constituyen propósitos razonables y deseables para sus acciones. Por ello, el principio de caridad incluye un supuesto fundamentalmente normativo. La intérprete no asumirá solamente que comparten creencias acerca de cómo es el mundo sino acerca de cómo *debería serlo*. La intérprete atribuirá al agente creencias descriptivas pero también valoraciones, con el objetivo de hacerlo inteligible.

La condición (iv) incluye dos puntos. En primer lugar, la intérprete debe asumir que el agente actúa, en general y excepto casos excepcionales de *akrasia* y auto-engaño, según su mejor juicio. En segundo lugar, la intérprete debe asumir que ellos comparten, en líneas generales, propósitos a largo plazo. En consecuencia, ella deberá asumir que ellos comparten la creencia de que ciertas acciones son deseables porque son medios para alcanzar ciertos fines, los cuales son medios para alcanzar ciertos otros fines, y así en adelante. También deberá la intérprete asumir que comparten creencias acerca de los fines a largo plazo. Naturalmente, la intérprete podrá estar en desacuerdo con algunos de estos medios y fines, pero deberá haber pasos en este encadenamiento de medios y fines (especialmente los más generales), en que ella se encontrará en acuerdo con el agente. Este acuerdo, asumido como tal por la intérprete, le hará posible comprender la cadena de medios y fines de menor generalidad. De esta manera, el principio de caridad tiene un doble elemento normativo. Por una parte, la intérprete no sólo atribuirá creencias sino también valoraciones. Por otra parte, la intérprete estará compelida a atribuir todos estos elementos imbricados para poder comprender al agente.

De esta manera, el holismo semántico contenido en el principio de caridad elimina las formas más radicales de relativismo e inconmensurabilidad a nivel epistemológico, pero también a nivel moral. Así como el discurso en general, también el discurso moral se halla gobernado por exigencias de consistencia y racionalidad. Pero tal vez la consecuencia más interesante sea una exigencia de solidaridad. Las condiciones de la interpretación nos exigen ver al otro como un semejante-diferente. Por lo diferente que es, nos cuestiona y se convierte en un reto permanente para nuestras propias creencias y deseos, así como nos obliga moralmente a intentar entenderlo, porque nuestra auto-comprensión está estrechamente ligada a la comprensión del otro. Del principio de caridad se deduce una obligación moral de tratar de entender al otro, de intentar incorporarlo a las posibilidades de nuestro propio discurso. Jeff Malpas lo dice de esta manera:

Uno podría ponerlo en un lenguaje ligeramente diferente y decir que una de las implicaciones éticas del holismo es que el cuidado por nosotros mismos es inseparable del cuidado por los otros. (Malpas 1992)

En efecto, el cuidado por uno mismo, como la auto-comprensión, pasa por el cuidado y comprensión del otro. La ética que se desprende del principio de caridad es una ética de la solidaridad y la tolerancia. Lejos de plantear el problema moral en términos de la pregunta por qué querría yo comprender al otro, para después intentar fundamentar una respuesta, la posición holista consideraría inevitable el desear comprender al otro y a uno mismo, mediante el reconocimiento de nuestro espacio compartido, como un solo proceso que es condición de posibilidad de toda acción en el mundo. A su vez, este proceso exige ver al otro, en palabras de Rorty, como “un compañero de travesía que sufre”.

Sin embargo, aquí reaparece el fantasma que persigue al principio de caridad desde su nacimiento: la posibilidad de que sea una técnica filosófica de justificación del ‘imperialismo psíquico’, como lo llama Robert Nozick (Nozick 1993: 53; cf. Lear 1991, MacIntyre 1989 y Hacking 1975). Aparentemente, esto se evidenciaría en la elección misma del término ‘caridad’, de posibles connotaciones paternalistas. Ahora no voy a desarrollar en profundidad este punto, que lo he discutido con detalle en otro lugar (cf. Quintanilla 2001), pero sí me gustaría sugerir que aunque no hay ningún vínculo explícito entre el principio de caridad y la virtud de la caridad, no puede ser tan arbitraria la elección del término. En efecto, la caridad es el cuidado desinteresado por el otro, pero es también el cuidado de uno mismo, y ambas formas de cuidado son inseparables. Se trata de una actitud de benevolencia en la interpretación y la interacción con el otro. Más que aludir a un tipo de paternalismo, caridad implica *philia*. De hecho, el latín *caritas*

es la traducción del griego *ágape*, que significa amor o cuidado, pero que además sugiere la idea de un espacio de enriquecimiento mutuo compartido, como lo fueron las comidas fraternales de los primeros cristianos.

Al hacer uso del principio de caridad, uno permite la comprensión del otro, pero también hace posible la comprensión de sí mismo, su auto-descripción y, por tanto, también su transformación. Al tratar de comprender al otro, la intérprete comparará sus propias creencias y deseos con aquellos atribuidos al agente; este proceso la obligará a poner en cuestión su propio sistema de estados mentales. Esto podrá conducir a una exigencia de tolerancia, adaptación y cambio, que permitirá la generación de un territorio común entre ambos interlocutores.

Así pues, interpretar al otro no es reconstruir sus contenidos mentales en los nuestros, como sostendría una concepción internalista de la mente, sino más bien es la actividad creativa de construir un territorio compartido: una comunidad de creencias, deseos, significados, valores, y objetos de la realidad. Este ámbito es elaborado pero también es hallado, porque pertenece al mundo que consideramos nos antecede en su objetividad. A su vez, es en este espacio donde justificamos nuestras más arraigadas convicciones, tanto en el terreno epistemológico como en el moral.

Bibliografía

- Astington, T., Harris, P., & Olson, D. (1988). *Developing Theories of Mind*. Cambridge: Cambridge University Press.
- Bennett, J. (1971). *Locke, Berkeley, Hume: Central themes*. Oxford: Clarendon Press.
- Bieri, P., Horstmann, R.P., & Kruger, L. (eds.). (1979). *Transcendental Arguments and Science*. Dordrecht: Reidel.
- Brandom, R. (1994). *Making it explicit. Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- Cavell, M. (1993). *The Psychoanalytic Mind. From Freud to Philosophy*. Cambridge Mass: Harvard University Press.
- (1998). "Triangulation, One's Own Mind and Objectivity". En: *International Journal of Psycho-Analysis* 79: 449-67.
- Davidson, D. (1980). "Mental events". En: *Essays on Actions and Events*. Oxford: Oxford University Press.
- (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
- (1984a). "Radical Interpretation". En: Davidson 1984.
- (1984b). "Belief and the Basis of Meaning". En: Davidson 1984.

- Davidson, D. (1984c). "Thought and Talk". En: Davidson 1984.
- (1984d). "On the Very Idea of a Conceptual Scheme". En: Davidson 1984.
- (1986). "A Coherence Theory of Truth and Knowledge". En: E. Lepore (ed.), *Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell.
- (1997). "Indeterminacy and anti-realism". En: C. Kulp (ed.), *Realism/Antirealism and Epistemology*. Lanham: Rowman & Littlefield. [También en Davidson 2001]
- (2001). *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press.
- Davies, M., / Stone, T. (1995a). *Folk Psychology. The Theory of Mind Debate*. Oxford: Blackwell.
- (1995b). *Mental Simulation. Evaluations and Applications*. Oxford: Blackwell.
- Devitt, M. (1981). *Designation*. Nueva York: Columbia University Press.
- Dilthey, W. (1913-1917). "Aufbau der geschichtlichen Welt in den Geisteswissenschaften". En: *Gesammelte Schriften*, vol. 5. Gottingen: Vandenhoeck & Ruprecht.
- Dreyfus, H. (1996). *Ser en el mundo. Comentarios a la división I de Ser y Tiempo*. Santiago de Chile: Cuatro Vientos.
- Grandy, R. (1973). "Reference, Meaning and Belief". En: *Journal of Philosophy* 70.
- Hacking, I. (1975). *Why does Language Matter to Philosophy*. Cambridge: Cambridge University Press.
- Hintikka, J. (1972). "Transcendental Arguments: Genuine and Spurious". En: *Nous* 6.
- Hume, D. (1975). *Inquiries Concerning Human Understanding and Concerning the Principles of Morals*, (reimpreso de la edición póstuma de 1777 y editado con introducción, tabla comparativa de contenido e índice analítico por L.A. Selby-Bigge, revisado por P.H. Nidditch). Oxford: Oxford University Press & Clarendon Press.
- Lear, J. (1991). *Love and its Place in Nature. A Philosophical Interpretation of Freudian Psychoanalysis*. New York: Noonday Press.
- Lewis, D. (1972). "Psychophysical and Theoretical Identifications". En: *Australasian Journal of Philosophy* 50: 249-58.
- (1983). "Radical interpretation". En: *Philosophical Papers*. Oxford: Oxford University Press.
- Locke, J. (1975). *An Essay concerning Human Understanding*. En: P. Nidditch (ed.). Oxford: Oxford University Press.
- Macdonald, G. / Pettit, P. (1981). *Semantics and Social Science*. Londres: Routledge & Kegan Paul.

- MacIntyre, A. (1989). "Relativism, Power and Philosophy". En: M. Krausz. (ed.), *Relativism: Interpretation and Confrontation*. Notre Dame, Indiana: University of Notre Dame Press.
- Malpas, J. (1990). "Transcendental Arguments and Conceptual Schemes. A Reconsideration of Körner's Uniqueness Argument". En: *Kant-Studien* 81: 232-51.
- (1992). *Donald Davidson and the Mirror of Meaning, Holism, Truth, Interpretation*. Cambridge: Cambridge University Press.
- Nozick, R. (1993). *The Nature of Rationality*. Princeton, New Jersey: Princeton University Press.
- Papineau, D. (1978). *For Science in Social Science*. Londres: McMillan.
- Quine, W.V. (1960). *Word and Object*. Cambridge MA: MIT Press.
- Quintanilla, P. (1997). "Teoría de la acción y racionalidad en Donald Davidson". En: *Revista Latinoamericana de Filosofía*, Vol. XXIII, 1.
- (1999). "La hermenéutica de Davidson: metáfora y creación conceptual". En: C. Caorsi (ed.), *Ensayos sobre Davidson*. Montevideo: Universidad de la República de Uruguay
- (2001). "El lugar de la racionalidad en la comprensión del otro". En: S. López, G. Portocarrero, R. Silva & V. Vich (eds.), *Estudios Culturales. Discursos, poderes, pulsiones*. Lima: Red para el desarrollo de las ciencias sociales en el Perú.
- Whiten, A. (ed.), (1991). *Natural Theories of Mind. Evolution, Development and Simulation of Everyday Mindreading*. Oxford: Basil Blackwell.
- Wilson, N.L. (1959). "Substances without substrata". En: *The Review of Metaphysics* 12.
- (1970). "Grice on meaning: the ultimate counter-example". En: *Nous* 4.
- Wittgenstein, L. (1972). *On Certainty*. Nueva York: Harper & Row.

Artículo recibido:xxxxxxx. Aceptado:xxxxx.

