



# Antropomorfismo en las interacciones entre humanos y robots: una conceptualización multidimensional

## Artículo traducido al español

**Título original:** Anthropomorphism in human–robot interactions: a multidimensional conceptualization  
<https://doi.org/10.1093/ct/qtac020>

## Rinaldo Kühne<sup>1</sup>

University of Amsterdam  
Amsterdam School of Communication Research  
<https://orcid.org/0000-0002-0411-1002>  
r.j.kuhne@uva.nl  
Holanda-Amsterdam

## Jochen Peter

University of Amsterdam  
Amsterdam School of Communication Research  
<https://orcid.org/0000-0002-2356-6619>  
J.Peter@uva.nl  
Holanda-Amsterdam

**Fecha de traducción:** 08/11/2023

## Resumen

Dado que los robots asumen cada vez más roles sociales (por ejemplo, asistentes, compañeros), el antropomorfismo (es decir, la cognición que posee una entidad, con características humanas) juega un papel destacado en la interacción humano-robot (IHR). Sin embargo, las conceptualizaciones actuales del antropomorfismo en la IHR no han distinguido adecuadamente entre precursores, consecuencias y dimensiones del antropomorfismo. Al construir y profundizar en investigaciones previas, conceptualizamos el antropomorfismo como una forma de cognición humana, que se centra en atribuir las capacidades mentales humanas a entidades no humanas como un robot. En consecuencia, las percepciones re-

lacionadas con la forma y el movimiento de un robot son precursoras potenciales del antropomorfismo, mientras que las atribuciones de personalidad y valor moral a un robot son consecuencias potenciales del antropomorfismo. A partir de la argumentación de que las conceptualizaciones multidimensionales reflejan mejor las facetas conceptuales del antropomorfismo, proponemos, con base en el marco de la Teoría de la Mente (ToM) de Wellman (1990), que el antropomorfismo en la IHR consiste en atribuir a un robot el pensamiento, el sentimiento, la percepción, el deseo y la elección. Concluimos discutiendo las aplicaciones de nuestra conceptualización en la investigación de la IHR.

## Palabras clave:

Interacción humano-robot (IHR); robots sociales; tecnología; Teoría de la Mente (ToM); agentes digitales

<sup>1</sup> Correspondencia del autor: Rinaldo Kühne, *Amsterdam School of Communication Research, ASCoR, University of Amsterdam*, P.O. Box 15791, 1001 NG, Amsterdam, The Netherlands. Email: r.j.kuhne@uva.nl.

©The Author(s) 2022. Published by Oxford University Press on behalf of International Communication Association.

Este es un artículo de Acceso abierto distribuido bajo los términos de la Licencia de Atribución *Creative Commons* (<https://creativecommons.org/licenses/by/4.0/>), que permite la reutilización, distribución y reproducción sin restricciones en cualquier medio, siempre que la obra original esté debidamente citada.





## Anthropomorphism in human–robot interactions: a multidimensional conceptualization

### Abstract

With robots increasingly assuming social roles (e.g., assistants, companions), anthropomorphism (i.e., the cognition that an entity possesses human characteristics) plays a prominent role in human–robot interactions (HRI). However, current conceptualizations of anthropomorphism in HRI have not adequately distinguished between precursors, consequences, and dimensions of anthropomorphism. Building and elaborating on previous research, we conceptualize anthropomorphism as a form of human cognition, which centers upon the attribution of human mental capacities to a robot. Accordingly,

perceptions related to a robot’s shape and movement are potential precursors of anthropomorphism, while attributions of personality and moral value to a robot are potential consequences of anthropomorphism. Arguing that multidimensional conceptualizations best reflect the conceptual facets of anthropomorphism, we propose, based on Wellman’s (1990) Theory-of-Mind (ToM) framework, that anthropomorphism in HRI consists of attributing thinking, feeling, perceiving, desiring, and choosing to a robot. We conclude by discussing applications of our conceptualization in HRI research.

### Keywords:

Human–robot interaction (HRI); social robots; technology; Theory-of-Mind (ToM); digital agents

## Introducción

La cognición que atribuye características humanas a una entidad –llamada antropomorfismo– como una respuesta psicológica fundamental que se basa en representaciones cognitivas típicamente adquiridas durante la infancia (Bjorklund y Causey, 2018). Si bien los humanos básicamente pueden antropomorfizar cualquier cosa (Epley *et al.*, 2007), las percepciones de la semejanza humana juegan un papel crucial en la interacción humano-robot (IHR) (Duffy, 2003; Fink, 2012; Kahn *et al.*, 2007) porque las características de los humanos a menudo se utilizan como principios rectores en el diseño de robots (Fong *et al.*, 2003). La comunicación humana, por ejemplo, se ha empleado a menudo como modelo para el diseño de tecnologías y se ha considerado clave para mejorar la usabilidad de los robots (Guzmán, 2018). En consecuencia, la investigación y el desarrollo en robótica se han centrado en hacer que los robots se comuniquen y se comporten más como humanos para facilitar su uso. Este desarrollo puede ser particularmente importante para los robots sociales (Breazeal, 2003). Estos no están dotados de características humanas para aumentar la facilidad en su uso, sino para crear interacciones y relaciones sociales significativas. (Breazeal, 2003; Fong *et al.*, 2003).

A pesar del papel crucial del diseño antropomórfico y, en consecuencia, de las percepciones de semejanza humana en la IHR, existen dos problemas con el concepto de antropomorfismo que actualmente obstaculizan el progreso en este campo. En primer lugar, el antropomorfismo no se ha distinguido sistemáticamente de los fenómenos psicológicos relacionados. En particular, no es evidente si ciertas cogniciones y conductas presentan precursores, consecuencias o elementos constitutivos del antropomorfismo. Por ejemplo, no está claro si las percepciones de la forma, el movimiento y el comportamiento de un robot conducen a su antropomorfización o si son un aspecto de esta.

Por otra parte, los marcos teóricos comúnmente utilizados en la IHR tampoco son concluyentes a este res-

pecto. Por ejemplo, en el marco de las “computadoras actores-sociales” (CASA), un enfoque fundamental en la investigación de la IHR, postula que la gente tiende a tratar las computadoras como si fueran personas reales. Las investigaciones han confirmado esta hipótesis al mostrar, por ejemplo, que los humanos son educados con computadoras o les aplican normas de reciprocidad a estas (Nass y Moon, 2000). Si bien el trabajo de CASA sugirió originalmente que las respuestas sociales a las computadoras son típicamente automáticas y sin sentido (Nass y Moon, 2000), estudios recientes sugieren que las respuestas sociales se ven facilitadas por representaciones informáticas más antropomórficas (Gong, 2008). En consecuencia, el papel del antropomorfismo en CASA parece merecer mayor atención. (Lee-Won *et al.*, 2020).

En segundo lugar, no existe consenso sobre la estructura dimensional del antropomorfismo en la IHR. Los académicos no están de acuerdo sobre si es mejor conceptualizar el antropomorfismo como unidimensional o multidimensional. Es decir, si se trata de un concepto monolítico (unidimensional) o si tiene múltiples facetas que constituyen el concepto general (multidimensional). Los defensores de una visión multidimensional, a su vez, no están de acuerdo sobre cuáles dimensiones debería incluir dicha conceptualización (Ruijten, 2018; Zhang *et al.*, 2008).

En ese sentido, por ejemplo, Epley *et al.* (2007) proponen que las atribuciones de pensamiento e intenciones conscientes son dimensiones cruciales del antropomorfismo. Sin embargo, los propios autores indican que facetas adicionales, como la atribución de emociones o características de comportamiento, también pueden ser importantes para definir el antropomorfismo. En consecuencia, los investigadores de la IHR han empleado diferentes conjuntos de criterios, así como dimensiones, al conceptualizar y operacionalizar el antropomorfismo (Bartneck *et al.*, 2009; Złotowski *et al.*, 2014).



Estas cuestiones exigen una visión general y una reflexión teórica, sobre cómo el antropomorfismo puede conceptualizarse clara y rigurosamente en la IHR. Sin tal conceptualización, los resultados de la investigación sobre el papel del antropomorfismo en la interacción entre humanos y robots seguirán siendo difíciles de comparar entre estudios y tanto la evidencia acumulada como el progreso en el campo se verán obstaculizados. A una escala más general, este artículo también responde al llamado de los académicos a una definición conceptual y operativa más sistemática y coherente de conceptos clave en este campo (por ejemplo, de Jong *et al.*, 2019).

Al abordar las cuestiones antes mencionadas, este artículo se basa en gran medida en trabajos previos sobre antropomorfismo de Epley y colegas (Epley *et al.*, 2007; Epley y Waytz, 2010) y en el marco de la Teoría de la Mente (ToM) de Wellman (1990). Se sostiene, en primer lugar, que los determinantes, dimensiones y consecuencias del antropomorfismo deben diferenciarse clara y consistentemente.

De forma específica, en este artículo se sostiene que las atribuciones de las capacidades mentales humanas están en el centro del antropomorfismo de los robots, mientras que las percepciones relacionadas con formas, movimientos y comportamientos son precursoras de este. Además, asignarle personalidad y valor moral a un robot debe considerarse consecuencia potencial de dicho antropomorfismo. No se aspira a crear una teoría completa sobre este asunto, la cual analice todos sus predictores, consecuencias y significado exacto. Más bien, el objetivo es aclarar qué facetas definitorias del antropomorfismo mencionadas con frecuencia deben considerarse predictores y consecuencias más que criterios constitutivos.

Asimismo, la conceptualización presentada en este trabajo es agnóstica en cuanto a la cuestión de la ética de los robots. Esta conceptualización no sugiere que se deba (o no) evaluar la ontología del funcionamiento de un robot antes de asignarle derechos. Más bien se centra exclusivamente en los mecanismos psicológicos que se

despliegan en la IHR, y no en cómo deben evaluarse los derechos de los robots (para una discusión detallada sobre la ética de los robots, ver Gunkel, 2012, 2022).

En segundo lugar, se postula que una conceptualización multidimensional es más adecuada que una conceptualización unidimensional para reflejar con precisión la complejidad conceptual del antropomorfismo. A partir del marco de ToM de Wellman (1990), se propone una conceptualización multidimensional del antropomorfismo del robot, que comprende el pensamiento, el sentimiento y la percepción. De esta forma, se propone una caracterización más focalizada y detallada y se distinguen cinco facetas. Se destaca que esta conceptualización puede ser aplicable a una amplia gama de interacciones entre humanos y robots y facilitar un análisis detallado de las relaciones del antropomorfismo con otros conceptos. Finalmente, se discuten las limitaciones de esta nueva conceptualización y sus implicaciones para una investigación futura.

## **Precusores y consecuencias del antropomorfismo**

La primera ambigüedad en la conceptualización del antropomorfismo en la interacción entre humanos y robots se refiere a los límites de esta noción, especialmente a la distinción entre dicho concepto, sus precursoros y consecuencias en este intercambio. Los estudiosos coinciden en que el antropomorfismo es una forma de cognición humana, que se enfoca en atribuir características humanas a una entidad no humana, lo que puede surgir cuando las personas observan o interactúan con un robot (Bartneck *et al.*, 2009; Duffy, 2003; Epley *et al.*, 2007; Fink, 2012).

Si bien el antropomorfismo se puede generar en respuesta a una amplia gama de robots: desde los que tienen apariencia de máquinas hasta los que tienen cualidades más humanas (Tan *et al.*, 2018), es probable que dotar a los robots de particularidades similares a las humanas facilite esta tendencia (Fink, 2012; Tan *et al.*, 2018). Por lo tanto, es necesario distinguir entre

antropomorfismo del diseño más humanizado de los robots y, más específicamente, de las propiedades de un robot que pueden (o no) provocar antropomorfismo (Fink, 2012; Ruijten, 2018; Yogeewaran *et al.*, 2016). Esta distinción se refleja en la investigación de la IHR que analiza cómo las características del robot que se espera que lo hagan más humano influyen en el grado en que los usuarios antropomorfizan un robot (Kiesler *et al.*, 2008; Salem *et al.*, 2013).

Sin embargo, los autores no están de acuerdo sobre dónde comienza y dónde termina el conocimiento de este concepto. Específicamente, no está claro cuáles saberes son precursores, cuáles lo constituyen y cuáles son consecuencias del antropomorfismo. Algunos autores han argumentado que este concepto incluye la percepción de que una entidad no humana tiene forma humana (Bartneck *et al.*, 2009; Liarokapis *et al.*, 2013; Waytz, Epley, *et al.*, 2010; Zhang y otros, 2008; Złotowski *et al.*, 2015). Esto comprende la percepción de que una entidad no humana tiene un tamaño humano (Zhang *et al.*, 2008), una morfología o forma humana (Liarokapis *et al.*, 2013; Zhang *et al.*, 2008) o un rostro humano (Bartneck *et al.*, 2009). La percepción de que una entidad no humana se mueve o se comporta como un humano también se ha considerado un aspecto del antropomorfismo (Bartneck *et al.*, 2009; Liarokapis *et al.*, 2013; Zhang *et al.*, 2008).

Con frecuencia, la atribución de capacidades mentales humanas se enfatiza en las definiciones de que se hacen sobre este fenómeno (Duffy, 2003; Epley *et al.*, 2007; Eyszel, 2017; Haslam y Loughnan, 2014). Desde este punto de vista, el antropomorfismo se refiere, en términos generales, a la creencia de que una entidad no humana puede pensar y sentir como un ser humano. Otras conceptualizaciones enfatizan que atribuir una personalidad (es decir, regularidades y consistencias en orientaciones y respuestas) (Higgins, 1990) a una entidad indica antropomorfismo (Kiesler y Goetz, 2002; Kim y Sundar, 2012; Salem *et al.*, 2013). Finalmente, Kahn *et al.* (2007) sugieren que la atribución de

valor moral intrínseco a un robot constituye un punto de referencia psicológico para determinar si el robot es percibido como humano. Esto implica que asignar valor moral a un robot es un aspecto adicional del concepto en cuestión. La definición se complica aún más ya que varias conceptualizaciones señalan varios de los aspectos anteriores como constituyentes del antropomorfismo (por ejemplo, Bartneck *et al.*, 2009; Kahn *et al.*, 2007).

Sostenemos que la atribución de capacidades mentales humanas a los robots presenta el núcleo del antropomorfismo robótico. Esta faceta conceptual se ha incluido en la mayoría de las discusiones sobre antropomorfismo en robótica (Broadbent, 2017; Duffy, 2003; Eyszel, 2017; Fink, 2012), aunque no en todas (por ejemplo, Zhang *et al.*, 2008), así como en conceptualizaciones psicológicas (Epley *et al.*, 2007), incluidas investigaciones sobre la deshumanización (Haslam y Loughnan, 2014), y en reflexiones filosóficas sobre la personalidad (Dennett, 1988). Esta definición más específica tiene, por tanto, la ventaja de que probablemente refleja un consenso académico sobre el significado del antropomorfismo.

Limitar el concepto de antropomorfismo robótico a la atribución de capacidades mentales humanas a los robots plantea la cuestión de qué papel desempeñan las facetas conceptuales excluidas (es decir, las percepciones de la forma y el movimiento humano de un robot, las atribuciones de personalidad y valor moral a un robot). En términos generales, sugerimos que estas facetas constituyen respuestas psicológicas que están relacionadas con el antropomorfismo, pero que aún son distintas. Más específicamente, dentro de la red nomológica del antropomorfismo en la IHR, las percepciones de la forma y el movimiento humano de un robot deben considerarse precursoras del antropomorfismo, mientras que las atribuciones de personalidad y valor moral a un robot deben considerarse consecuencias del antropomorfismo.



Como Epley *et al.* (2007), el antropomorfismo es una “inferencia sobre características no observables de un agente no humano, en lugar de informes descriptivos del comportamiento observable o imaginado de un agente no humano” (p. 865). Por tanto, las percepciones de la forma y el movimiento del robot pueden distinguirse del antropomorfismo porque se refieren a sus características observables. Este punto de vista resuena con la posición de Eysel (2017), quien argumentó, al referirse a Epley *et al.* (2007), que el antropomorfismo en la IHR va “más allá de simples juicios de una entidad con respecto a su forma antropomórfica o su apariencia de vida” (p. 366).

Las investigaciones indican que las percepciones de las características observables de los robots influyen en los juicios sobre las capacidades mentales de los robots. Por ejemplo, Barco *et al.* (2020) y Hegel *et al.* (2008) demostraron que los robots con forma humana están más antropomorfizados que los robots funcionales y zoomórficos. Además, se ha demostrado que la percepción de que un objetivo no humano se mueve a una velocidad similar a la del movimiento humano aumenta la atribución mental (Morewedge *et al.*, 2007). Como analizan Urquiza Haas y Kotrschal (2015) en su revisión, la investigación neurológica corrobora que el “centro de procesamiento de información social (...) [es activado por] (...) el movimiento biológico, el movimiento del cuerpo humano, las manos y la boca, del cerebro humano” (p. 169). De manera similar, se ha demostrado que las conductas comunicativas y, en particular el uso del habla, aumentan la probabilidad de que una entidad sea juzgada como humana (Schroeder y Epley, 2016). En resumen, es plausible considerar las percepciones relacionadas con características observables de los robots, como la forma y el movimiento, como cogniciones “ascendentes” (o precursoras) que pueden resultar en antropomorfismo robótico.

También se requiere una demarcación conceptual más clara en el lado “descendente” –de las consecuen-

cias– del antropomorfismo. En particular, este concepto debe distinguirse de la atribución de personalidad y valor moral a un robot, dado que esto se ha incluido previamente en conceptualizaciones del antropomorfismo. Tanto las explicaciones filosóficas (Dennett, 1987, 1988) como las psicológicas (Epley y Waytz, 2010) sobre el antropomorfismo sugieren que las atribuciones de personalidad y valor moral a un robot deben considerarse consecuencias más que dimensiones de su antropomorfismo.

Dennett (1987, 1988) propuso que los humanos pueden utilizar tres estrategias diferentes para analizar y predecir su entorno. Junto a la postura física (analizar el entorno en términos de propiedades y leyes físicas) y la postura de diseño (analizar las entidades en términos de sus funciones o propósitos previstos), este autor propone lo que él llama la postura intencional, que es la más central de este artículo. En ella se enfoca principalmente en el antropomorfismo (es decir, la atribución de capacidades mentales humanas), ya que implica que las personas pueden tratar a otras entidades como “agentes racionales” (Dennett, 1987, p. 17) para predecir su comportamiento. Esta explicación es la más útil para comprender y predecir el comportamiento de otros humanos, pero también se aplica a entidades no humanas, como animales y fenómenos naturales. Como señala el mencionado autor (1987), adoptar la postura intencional es una puerta de entrada para cogniciones “sociales” posteriores:

“Primero se decide tratar el objeto cuyo comportamiento se va a predecir como un agente racional; luego descubres qué creencias debería tener ese agente, dado su lugar en el mundo y su propósito. Luego se determina qué deseos debería tener, basándose en las mismas consideraciones, y finalmente se predice que este agente racional actuará para promover sus objetivos a la luz de sus creencias” (p. 17).

Una visión similar surge de Epley y Waytz (2010), quienes sostienen que el antropomorfismo es un precursor de las cogniciones sobre los estados mentales y

comportamientos de otras personas. Asimismo, Margolin (2021) propone que los individuos busquen información sobre las “tendencias de los agentes en la toma de decisiones” cuando intentan comprender su comportamiento (p. 718). En línea con estos puntos de vista, la investigación neurológica muestra una “distinción neurológica en el procesamiento de los aspectos físicos y sociales del mundo” (Urquiza-Haas y Kotrschal, 2015: p. 169) y que las percepciones de que una entidad es humana pueden estimular “circuitos cerebrales sociales” (Hortensius y Cross, 2018). En consecuencia, tiene sentido considerar el antropomorfismo de los robots como un precursor de la atribución de una personalidad específica a un robot: los humanos primero deciden si puede ser considerado como humano y, si este es el caso, pueden evaluarlo según sus relaciones con las personas.

Varios autores también sugieren que el antropomorfismo precede a la atribución de un valor moral a una entidad. Haslam y Loughnan (2014) y Waytz, Gray, *et al.* (2010), por ejemplo, sugieren que el valor moral y la responsabilidad moral se atribuyen típicamente a entidades que son capaces de actuar y experimentar, dos capacidades mentales esenciales de los seres humanos. La agencia (incluida la intencionalidad y la toma de decisiones autónoma) es crucial para la asignación de responsabilidad moral, mientras que la capacidad de experimentar felicidad y dolor es esencial para la asignación de valores y derechos morales. En consecuencia, antropomorfizar al robot influye en si es considerado un agente moral (Sullins, 2006). Con base en la literatura, es plausible distinguir el antropomorfismo (la atribución de capacidades mentales humanas a una entidad no humana) de las cogniciones anteriores (es decir, percepciones de forma y movimiento como humanas), así como de los conocimientos posteriores que se derivan de él (es decir, atribuciones de personalidad y valor moral).

En comparación con otras conceptualizaciones del antropomorfismo, la que se desarrolla en este ar-

tículo está más centrada, lo cual es ventajoso por al menos dos razones. Primero, es más “adecuado”, lo que aumenta su valor heurístico y facilita futuras investigaciones en la IHR. Una definición adecuada no es ni demasiado estrecha ni demasiado amplia en su significado (Westermann, 2000). Una conceptualización amplia del antropomorfismo corre el riesgo de subsumir diferentes tipos de conceptualizaciones, que potencialmente se influyen entre sí y forman un mecanismo causal, bajo un solo concepto. Esto puede obstaculizar la investigación sobre la interacción de conceptos que merecen atención por derecho propio. Por el contrario, distinguir este concepto de sus posibles precursores y consecuencias posibilita un análisis causal teóricamente sólido de las relaciones del antropomorfismo con nociones relacionadas, no solo con las mencionadas anteriormente, sino con una amplia gama de precursores potenciales (por ejemplo, conocimiento del agente provocado, motivación de la efectividad; como puede verse en Epley *et al.*, 2007) y consecuencias (por ejemplo, la aceptación de robots; ver Duffy, 2003) que se discuten en la investigación de la IHR. Nuestra conceptualización más enfocada no solo facilita distinciones “verticales” a lo largo de la cadena causal, sino también distinciones “horizontales” entre antropomorfismo y conceptos similares como animicidad (Scholl y Tremoulet, 2000) y presencia social (Biocca *et al.*, 2003). Sin embargo, una discusión en profundidad de estas distinciones está más allá del alcance de este artículo porque requiere un análisis conceptual elaborado de cada concepto similar.

En segundo lugar, una definición más centrada del antropomorfismo apoya una conceptualización “estable”, que reemplaza las definiciones dependientes del contexto. Se dice que es estable o duradera si puede utilizarse en diversas condiciones (McLeod y Pan, 2005). El antropomorfismo puede ocurrir en respuesta a una amplia gama de entidades (por ejemplo, objetos, animales, artefactos, fenómenos naturales). En la IHR, más específicamente, puede surgir en respuesta a una gran variedad de robots, incluidos los antropomórficos,



zoomórficos, caricaturizados y funcionales, así como los encarnados y no encarnados (Fong *et al.*, 2003).

Por tanto, una definición amplia de este concepto corre el riesgo de estipular criterios de antropomorfismo que solo son relevantes para un subconjunto de robots, con el resultado de que las respuestas antropomórficas no se pueden comparar entre modelos de robots. Por ejemplo, si la forma humana es un criterio (que lo sería en una definición amplia), los robots con y sin cuerpo difícilmente pueden compararse en su grado de antropomorfismo. Esto no significa que las características de los robots y las percepciones de dichas características deban descartarse porque solo sean relevantes en un subconjunto de contextos de la IHR. Más bien, se sugiere que estas características del robot y las percepciones correspondientes son predictoras del antropomorfismo que dependen del contexto. Tal como se define acá, por el contrario, el antropomorfismo es un fenómeno psicológico más general que puede surgir en una amplia gama de contextos de la IHR. Por lo tanto, una definición más centrada da cuenta de la generalidad del fenómeno, al tiempo que permite el estudio significativo de predictores dependientes del contexto que lo provocan.

### Conceptualizaciones unidimensionales frente a las multidimensionales del antropomorfismo

Como se mencionó anteriormente, los académicos de la IHR no están completamente de acuerdo, en primer lugar, sobre si es mejor conceptualizar el antropomorfismo como unidimensional o multidimensional y, en segundo lugar, qué dimensiones debe incluir una conceptualización multidimensional (Ruijten, 2018; Zhang *et al.*, 2008). Entender un concepto como unidimensional implica que sus facetas se superponen en gran medida en su significado y, en consecuencia, que sus indicadores, estadísticamente hablando, dependen de un factor. Una visión multidimensional implica que las facetas del concepto tienen un significado (parcialmente) distinto y, en consecuencia, que sus indicadores dependen de múltiples factores.

Con frecuencia se emplean conceptualizaciones unidimensionales del antropomorfismo (Barco *et al.*, 2020; Salem *et al.*, 2013; Bartneck *et al.*, 2009), dentro de las cuales destaca esta última. Las definiciones unidimensionales no son problemáticas per se si están arraigadas en una definición teórica precisa. Sin embargo, se parte de la idea de que es preferible una conceptualización multidimensional por dos razones. Primero, los humanos poseen una variedad de capacidades mentales (por ejemplo, sentir y pensar) y es probable que hagan referencia a múltiples capacidades mentales y las diferencien entre sí al evaluar si un robot es humano. Por tanto, es plausible suponer que el antropomorfismo de los robots incluye facetas múltiples y distintas.

En segundo lugar, una conceptualización unidimensional, que no permite facetas múltiples y distintas, corre el riesgo de ser demasiado abstracta e imprecisa. En consecuencia, las conceptualizaciones unidimensionales han provocado definiciones operativas discrepantes. Por ejemplo, Bartneck *et al.* (2009) utilizaron indicadores como *humanlike*, “realista” y “natural” para formar una medida unidimensional del antropomorfismo. Waytz, Morewedge *et al.* (2010) emplearon una serie de operacionalizaciones unidimensionales diferentes, incluida la medida de un solo elemento que la computadora parece “comportarse como si tuviera sus propias creencias y deseos” (p. 414) y una medida unidimensional que consta de los tres indicadores de que un dispositivo tiene “mente propia”, “intenciones, libre albedrío, conciencia” y que parece “experimentar emociones” (p. 415). Una conceptualización multidimensional, por el contrario, puede aumentar la precisión conceptual. Al especificar subdimensiones más concretas y, por lo tanto, apoya una definición operativa más consistente y comparable de antropomorfismo.

Una conceptualización multidimensional frecuentemente empleada tiene sus raíces en el modelo dual de deshumanización de Haslam, que aborda porque las personas niegan humanidad a otras personas o gru-

pos sociales (Haslam, 2006; Haslam *et al.*, 2009; Haslam y Loughnan, 2014). El modelo diferencia dos categorías de capacidades mentales que pueden atribuirse o no a otros: características exclusivamente humanas, que distinguen a los humanos de los animales y entidades no humanas (por ejemplo, refinamiento), y características que corresponden esencial o típicamente a la naturaleza humana (por ejemplo, calidez interpersonal). La deshumanización ocurre cuando las personas perciben que los demás carecen de estas características (Haslam y Loughnan, 2014).

El modelo de deshumanización de Haslam (Haslam, 2006; Haslam y Loughnan, 2014) es una base potencialmente útil para la conceptualización del antropomorfismo en la IHR porque describe criterios mentales que los individuos pueden usar al evaluar la semejanza humana de un robot. En consecuencia, los investigadores de la IHR han postulado que el antropomorfismo de los robots incluye una dimensión “exclusivamente humana” y una “naturaleza humana” (Eyssel *et al.*, 2010; Złotowski *et al.*, 2014) y varios estudios de la IHR hacen referencia a esta distinción al medir el antropomorfismo (Eyssel *et al.*, 2011; Ferrari *et al.*, 2016; Salem *et al.*, 2013).

Sin embargo, las características exclusivamente humanas constituyen en sí mismas categorías mentales amplias, que abarcan una variedad de conceptos psicológicos. Según Haslam (2006), la singularidad humana se refiere a la civilidad, el refinamiento, la sensibilidad moral, la racionalidad y la lógica y la madurez. La naturaleza humana se refiere a la capacidad de respuesta emocional, la calidez interpersonal, la apertura cognitiva, la agencia, la individualidad y la profundidad. Debido a esta amplitud conceptual, distinguir las características exclusivamente humanas como dimensiones constitutivas del antropomorfismo no aclara suficientemente su significado.

Una segunda conceptualización multidimensional del antropomorfismo tiene sus raíces en el trabajo de Gray, Gray y Wegner (2007) sobre la percepción mental. Proponen que las personas consideren dos facultades

mentales al evaluar si otras entidades tienen mente: si la entidad puede actuar y decidir de forma autónoma (agencia) y si es sensible (experiencia). En consecuencia, los académicos de la IHR han argumentado que el antropomorfismo de los robots es bidimensional que consta de agencia y experiencia (Hortensius y Cross, 2018; Złotowski *et al.*, 2015). Los académicos han reconocido una superposición conceptual entre los marcos de Haslam y Gray *et al.* con características exclusivamente humanas que se mapean ampliamente en la agencia y las características de la naturaleza humana en la experiencia (Hortensius y Cross, 2018; Złotowski *et al.*, 2015). Sin embargo, al igual que la distinción entre unicidad humana y naturaleza humana (Haslam, 2006), la distinción agencia-experiencia da como resultado subdimensiones del antropomorfismo que son conceptualmente confusas. Además, si bien la distinción agencia-experiencia ha influido en las operacionalizaciones del antropomorfismo y la percepción mental en la IHR (por ejemplo, Ferrari *et al.*, 2016; Fraune, 2020; Fraune *et al.*, 2020; Trovato y Eyssele, 2017), aún no se ha convertido en un estándar teórico establecido en esta interacción.

### Uso del marco ToM para especificar dimensiones del antropomorfismo

Actualmente, tanto las conceptualizaciones unidimensionales como las multidimensionales del antropomorfismo en la IHR son ambiguas: las conceptualizaciones unidimensionales no distinguen subdimensiones relevantes del antropomorfismo y las conceptualizaciones multidimensionales proponen subdimensiones que no son del todo claras en sí mismas. En consecuencia, el antropomorfismo se ha operacionalizado de manera inconsistente en la IHR. En este contexto, proponemos que un conjunto de dimensiones centrales del antropomorfismo, que explican y complementan las principales en los marcos existentes (Haslam y Loughnan, 2014; Gray *et al.*, 2007), pueden basarse en el marco de ToM. El principal supuesto del marco de ToM es que los humanos poseen una teoría ingenua de la mente, que propo-



ne una serie de facultades mentales que poseen las personas. (por ejemplo, sentir y pensar) y que puede usarse para predecir el comportamiento de los seres humanos (Bjorklund y Causey, 2018).

El marco de ToM se ha empleado principalmente en investigaciones sobre el desarrollo que investiga a qué edad los niños adquieren aspectos específicos de una ToM (Bjorklund y Causey, 2018). El marco también se ha mencionado en el contexto de la IHR. En estudios orientados al diseño, los académicos han postulado que los robots deberían estar equipados con una ToM para facilitar las interacciones sociales adecuadas entre humanos y robots (Benninghoff *et al.*, 2013; Scassellati, 2002). En la investigación centrada en el usuario, algunos académicos han notado una relación conceptual entre ToM y antropomorfismo. Stafford *et al.* (2014) postularon que “la teoría de la percepción mental está relacionada con el antropomorfismo, en el sentido de que las personas atribuyen capacidades mentales a personajes no humanos” (p. 20). Airenti (2018, p. 10) expresó una opinión similar: “En la atribución antropomórfica, los niños usan la misma teoría de las habilidades mentales que utilizan en interacciones con humanos”.

Este autor descubrió que la semejanza humana efectivamente aumentaba el uso de ToM con robots. Más allá de la investigación de la IHR, Epley y Waytz (2010) señalan que la percepción mental es un proceso inferencial que “normalmente se denomina teoría de la mente” (p. 505). Los relatos anteriores indican que emplear ToM en interacciones humano-humano (HHI) y antropomorfizar robots en IHR implica procesos relacionados, excepto que, en HHI, el otro es un humano y, en IHR, el otro es un robot. Sin embargo, existe una diferencia adicional entre HHI y IHR, que es particularmente relevante en el contexto del antropomorfismo en IHR y es que este precede a la aplicación de la ToM, es decir, la explicación del comportamiento del robot sobre la base de capacidades mentales. Aplicar ToM para comprender el comportamiento de un robot requiere que este sea primero (al menos en parte) antropomorfizado o, en la ter-

minología de Dennett (1987, 1988), que se adopte una postura intencional hacia el robot. Según Dennett (1987), primero hay que decidir tratar un objeto como un agente racional, antes de que se le puedan atribuir creencias y deseos.

En consecuencia, Epley (2018) sostiene que “en lugar de tratar a otras personas como objetos, en cambio, atribuimos una mente a otra persona, completa con conceptos como intenciones, deseos, actitudes y creencias que pueden usarse para explicar su comportamiento” (pp. 592-593). Si bien las evaluaciones iniciales de las facultades mentales de los demás son relevantes en las interacciones humanas (por ejemplo, en interacciones con niños o personas con discapacidades mentales), son esenciales en IHR porque los robots (sociales) difieren ampliamente en sus capacidades y en su similitud con los humanos (Fong *et al.*, 2003).

El hecho de que el antropomorfismo preceda a la aplicación de la ToM aclara por qué este y la ToM implican subdimensiones correspondientes: la aplicación de la ToM a un robot presupone lógicamente que este posee un conjunto específico de facultades mentales humanas: solo si el robot tiene estas facultades pueden basarse en ellas las explicaciones de su comportamiento. El antropomorfismo constituye así una cognición en la que un individuo evalúa si el robot posee las respectivas facultades mentales humanas y en qué medida. Cada evaluación de una facultad mental distinta constituye, a su vez, una dimensión del antropomorfismo. En consecuencia, una inspección de los principales elementos de la ToM de las personas debería revelar las principales dimensiones del antropomorfismo.

El modelo de ToM de Wellman (1990; véase también Wellman y Bartsch, 1988) es particularmente adecuado para derivar las principales dimensiones del antropomorfismo porque analiza explícitamente las facultades mentales que comprende la teoría de la mente de las personas. En el centro del modelo está la noción de razonamiento de creencia-deseo: se supo-

ne que las acciones de las personas están arraigadas en sus creencias y deseos.

Las creencias constituyen el conocimiento de un individuo sobre el estado del mundo. Los deseos son lo que el individuo quiere o espera. Se supone que ambos predicen las intenciones de las personas y reflejan lo que pretenden o planean hacer. Las intenciones, a su vez, se consideran el principal predictor de acciones: “Las intenciones son planes para hacer realidad ciertos deseos; las intenciones incorporan diversas creencias a las acciones” (Wellman, 1990, p. 111).

Las creencias y deseos de los individuos están influenciados por percepciones y emociones básicas y fisiológicos: se supone que las percepciones (como, ver, oír) informan a otros individuos sobre el mundo externo y forman la base de las creencias, mientras que las emociones y la fisiología básicas (por ejemplo, el hambre, el miedo) funcionan como un sistema interno, fuente de información que predice los deseos.

El modelo de ToM de Wellman (1990) comprende dos conceptos adicionales: emociones cognitivas y sensaciones. Sin embargo, ambos términos se superponen en gran medida con los discutidos anteriormente: las emociones cognitivas son un “concepto intermedio entre la emoción y el pensamiento” (Wellman, 1990, p. 114), y las sensaciones pueden ubicarse entre la percepción y la fisiología (el papel de las sensaciones dentro del modelo no se explica en detalle). Por tanto, las dos nociones no parecen constituir facultades mentales distintas que puedan informar dimensiones del antropomorfismo.

Al partir del modelo de ToM de Wellman (1990), se propone que el antropomorfismo de los robots incluye cinco dimensiones que reflejan los principales componentes de la ToM de las personas. Los individuos atribuyen estas capacidades a los robots de una manera no binaria: evalúan hasta qué punto un robot posee esas capacidades.

La primera dimensión es la atribución del pensamiento a un robot. Se utiliza el término pensamiento en lugar de creencias porque, como señala Wellman (1990), las creencias pertenecen al concepto más amplio de pensamiento y son producidas por procesos de pensamiento, como el razonamiento, el aprendizaje y la memoria. En consecuencia, se define el pensamiento como todos los procesos cognitivos de nivel superior (Blanchette y Richards, 2010) o todos los tipos de procesamiento de información “conscientes y con esfuerzo” (Smith y DeCoster, 2000, p. 108) que implican “la recuperación intencional de información explícita, reglas representadas simbólicamente (...) y su uso para guiar el procesamiento” (Smith y DeCoster, 2000, p. 108). Cuando se atribuye el pensamiento a un robot, se cree que el hardware y el software del robot permiten algunas formas de procesos cognitivos superiores. El pensamiento es parte de conceptualizaciones previas del antropomorfismo y la percepción mental: refleja aspectos de características exclusivamente humanas en la teoría de la deshumanización (Haslam, 2006; Haslam *et al.*, 2009) y de agencia en la teoría de la mente de Gray *et al.* (2007) percepción. La atribución del pensamiento (o procesos cognitivos de nivel superior) también se ha identificado previamente como un aspecto del antropomorfismo en IHR (Duffy, 2003; Fussell *et al.*, 2008; Trovato y Eyssele, 2017).

Los conceptos de fisiología, emoción y sensación en el modelo de Wellman (1990) se resumen en la segunda dimensión del antropomorfismo: la atribución de sentimiento a un robot. Los sentimientos son las experiencias individuales de cambios y reacciones corporales (LeDoux, 1999), como el hambre, la excitación sexual y el dolor. Un aspecto importante del sentimiento es la experiencia de emociones como la ira, el miedo y la felicidad. Se sigue a Leventhal y Scherer (1987) quienes proponen que los sentimientos pueden ser provocados por procesos sensoriomotores, esquemáticos y conceptuales.

Los sentimientos pueden ser producidos por procesos que van desde lo automático y lo reflejo, que se corresponde con el concepto de sensaciones de Wellman (1990), hasta lo volitivo y reflexivo, que se corresponde con el concepto de emociones cognitivas de Wellman. Atribuir sentimientos a un robot no supone que este posea una fisiología humana. De manera similar a la atribución del pensamiento a un robot, refleja más bien la creencia de que el *hardware* y el *software* de un robot permiten experiencias subjetivas. La dimensión emocional del antropomorfismo refleja un aspecto de la naturaleza humana en la teoría de la deshumanización, es decir, emocionalidad (Haslam, 2006; Haslam *et al.*, 2009). Se corresponde estrechamente con la dimensión de la experiencia en la concepción de percepción mental de Gray *et al.* (2007). Además, varios académicos de la IHR han mencionado la atribución de emociones como un aspecto del antropomorfismo (Duffy, 2003; Fussell *et al.*, 2008; Lemaignan *et al.*, 2014; Wang y Krumhuber, 2018).

La tercera dimensión del antropomorfismo es la atribución de percepción a un robot, lo que significa que este es capaz de sentir su entorno (en particular, visión, oído, tacto, gusto y olfato) e interpreta esta información. Una vez más, atribuir la percepción a un robot no supone una fisiología humana porque los sensores pueden emular los sentidos humanos. Subsumimos la sensación bajo sentimiento en lugar de percepción porque la definición de Wellman se refiere a experiencias internas como mareos, náuseas y dolor. En consecuencia, postulamos que los sentimientos se refieren a la experiencia de estados internos, mientras que las percepciones se refieren a la experiencia de estímulos externos.

Los académicos de la IHR no suelen identificar la percepción como una dimensión distinta del antropomorfismo. Sin embargo, es relevante porque las percepciones son un aspecto esencial de la ToM. El marco y la aplicación de ToM a un robot suponen que este tiene al menos un acceso mínimo a su entorno. Además, los académicos han propuesto que los agentes artificiales, como los robots, deberían poder

percibir su entorno para funcionar y ser considerados entidades sociales. Hubbard (2011) señala que "para ser una entidad viva de cualquier tipo, uno debe tener la capacidad de interactuar significativamente con el entorno recibiendo y decodificando entradas de su entorno y enviándole datos inteligibles" (p. 419). La investigación sobre robots sociales (es decir, los que emulan la sociabilidad con el fin de permitir relaciones sociales con los humanos) indica que los estos solo pueden cumplir funciones sociales si pueden percibir su entorno (social) (Breazeal, 2003; Fong *et al.*, 2003).

La cuarta dimensión es la atribución de desear o querer a un robot. A esta facultad también se le ha denominado motivación (Baumeister, 2016; Ryan y Deci, 2000). Según Baumeister (2016), "es una condición de un organismo que incluye un sentido subjetivo (no necesariamente consciente) de desear algún cambio en uno mismo y/o medio ambiente" (p. 1). Atribuir deseo a un robot corresponde a creer que este posee necesidades y preferencias que desea gratificar. Desear (o querer) está estrechamente relacionado con la intención, como se indicó anteriormente y según lo elaborado por Wellman (1990, p. 110): "Uno puede tener un deseo (un deseo, una esperanza, una fantasía) hacer algo, pero, a pesar de ello, nunca tiene la intención (planea, apunta, decide) de hacerlo".

Los individuos pueden poseer una variedad de necesidades, pero pueden elegir cuáles intentan satisfacer y a través de qué medios. En consecuencia, los comportamientos de los individuos no están completamente determinados por sus deseos (y creencias) porque pueden elegir un curso de acción. Se han utilizado diferentes términos para describir esta idea en el contexto de agentes artificiales, incluida la intencionalidad (Bigman *et al.*, 2019), el libre albedrío (Krausová y Hazan, 2013), la agencia (Himma, 2009) y la autonomía (Beer *et al.*, 2014). Usamos el término elección para describir la capacidad de elegir libremente entre múltiples cursos de acción, que es la quinta y última dimensión del antropomorfismo de los robots.

Tanto el deseo como la elección constituyen características de la naturaleza humana en la teoría de la deshumanización (Haslam, 2006; Haslam *et al.*, 2009). Además, el deseo refleja un aspecto de la experiencia y la elección, un aspecto de la agencia en la teoría de la percepción mental de Gray *et al.* (2007). Sin embargo, al igual que la percepción, el deseo y la elección no suelen identificarse como dimensiones distintas del antropomorfismo en la IHR.

En resumen, de acuerdo con investigaciones previas y especialmente el marco de ToM de Wellman (1990), se propone que la antropomorfización de los robots consiste en atribuir cinco capacidades mentales humanas a un robot: pensar, sentir, percibir, desear y elegir. Dichas capacidades mentales constituyen antropomorfismo robótico porque son parte integral de la explicación y predicción del comportamiento humano.

En consecuencia, si alguna de las dimensiones no se incluyera en una definición de antropomorfismo robótico faltaría un punto de referencia crucial sobre la semejanza humana de una mente. Como la conceptualización que se presenta en este artículo explica las principales dimensiones del antropomorfismo representa una especificación de la postura intencional de Dennett (1987) y las ideas existentes del antropomorfismo (por ejemplo, Epley *et al.*, 2007). Al especificar las principales dimensiones del antropomorfismo, en este artículo se mejora la precisión de la definición y disminuye la ambigüedad (McLeod y Pan, 2005). La representación del antropomorfismo se ilustra en la Figura N°1, mientras que la Tabla N°1 detalla las ventajas de la perspectiva que se desarrolla en este artículo respecto de enfoques anteriores. En la tabla se enumeran tres criterios principales de calidad de las definiciones (estabilidad, adecuación y precisión) y especifica: (a) problemas de conceptualizaciones existentes relacionadas con estos criterios; (b) cómo la conceptualización acá presentada aborda estos problemas; (c) y las ventajas de esta para futuras investigaciones.

## Investigaciones futuras y limitaciones

La conceptualización del antropomorfismo que se hace en este texto puede impulsar la investigación en este campo y abrir múltiples direcciones para futuras investigaciones. En primer lugar, demarcar claramente los límites conceptuales del antropomorfismo simplifica la exploración de sus antecedentes y repercusiones. Este tipo de investigación no se limita al estudio de conceptos tradicionalmente asociados al antropomorfismo. Más bien, la perspectiva conceptual desarrollada acá abre la puerta al estudio de diversos factores predictivos potenciales como la interacción con robots y las características del usuario, así como sus posibles consecuencias como la aceptación del robot y la formación de relaciones con ellos. Además, debido a que tal conceptualización es multidimensional, permite investigar si dimensiones específicas del antropomorfismo responden de manera diferenciada a los factores predictivos, generan resultados diversos y cómo se relacionan entre sí estas dimensiones.

En segundo lugar, esta conceptualización permite establecer distinciones claras entre el antropomorfismo y conceptos afines que se estudian con frecuencia en la IHR, como la animacidad (Scholl y Tremoulet, 2000) y la presencia social (Biocca *et al.*, 2003). Ello estimula la investigación sobre los roles diferenciales de estos conceptos en IHR. Estas distinciones ofrecen la posibilidad de investigar, por ejemplo, los efectos específicos de la morfología del robot en el antropomorfismo, la animacidad y la presencia social (Barco *et al.*, 2020).

En tercer lugar, esta perspectiva del antropomorfismo excluye las cogniciones ascendentes dependientes del contexto, en particular, las percepciones de las formas, movimientos y comportamientos de los robots. Por tanto, puede aplicarse a una amplia gama de contextos de IHR, que abarcan diferentes tipos de robots desde antropomórficos hasta zoomorfos, caricaturizados y funcionales (Fong *et al.*, 2003). Esta aplicación in-



cluye robots con un diseño altamente antropomórfico (androides), así como unos diseños mínimamente antropomórficos (robots industriales) en diversos escenarios de interacción (como juegos o colaboración con robots). Al mismo tiempo, la conceptualización desarrollada puede ser complementada con cogniciones ascendentes dependientes del contexto para explorar procesos psicológicos idiosincrásicos. Por ejemplo, al estudiar robots humanoides, se puede investigar si la percepción de que el robot tiene una forma humana influye en el antropomorfismo.

Existen al menos tres limitaciones en esta conceptualización. Primero, supone que el antropomorfismo de los robots comprende cogniciones explícitas a las que los individuos pueden acceder conscientemente; por lo tanto, no cubre el antropomorfismo implícito (Urquiza-Haas y Kotrschal, 2015) o inconsciente (Kim y Sundar, 2012; Nass y Moon). Sin embargo, la perspectiva que se muestra en este artículo es compatible con la noción de que los procesos automáticos e inconscientes, así como los reflexivos y conscientes, pueden provocar antropomorfismo (Echterhoff *et al.*, 2006; Kim y Sundar, 2012). En consecuencia, puede usarse para complementar el marco CASA, que postula que las respuestas sociales a las máquinas resultan de la activación inconsciente de guiones sociales (Nass y Moon, 2000).

En línea con desarrollos más amplios recientes del marco CASA (Lombard y Xu, 2021), nuestra conceptualización se suma al especificar una ruta adicional (la atribución consciente o explícita de las capacidades mentales humanas) a través de la cual los robots pueden provocar el uso de las redes sociales, reglas y heurísticas. Esta ruta explícita puede volverse cada vez más importante en el futuro a medida que los robots adquieran sucesivamente características más parecidas a las humanas. De hecho, acá se sugiere que futuras investigaciones aborden cómo los procesos conscientes y sin sentido producen antropomorfismo. Esto es particularmente relevante ya que los

procesos conscientes y sin sentido pueden fomentar resultados diferentes sobre la semejanza humana de un robot (Banks, 2020; Echterhoff *et al.*, 2006).

En segundo lugar, esta conceptualización del antropomorfismo aún no se ha traducido en un instrumento de medición. Según el contexto de aplicación, pueden ser más adecuados diferentes modelos de medición: uno reflexivo implica que la variación del antropomorfismo crea variación en sus indicadores, que se supone que están altamente correlacionados (Coltman *et al.*, 2008). Un modelo de este tipo puede adecuarse en situaciones en las que los individuos hacen juicios rápidos e intuitivos sobre la semejanza humana de un robot (por ejemplo, cuando se evalúan robots que son muy parecidos a los humanos o que no se parecen en nada a los humanos). Por el contrario, un modelo formativo implica que la variación en los indicadores, que no se supone que estén altamente correlacionados, crea variación en el antropomorfismo (Coltman *et al.*, 2008). Un modelo de esta clase puede usarse cuando los individuos analizan gradualmente las capacidades de un robot (por ejemplo, al evaluar casos límite).

**Figura N° 1.** Conceptualización del antropomorfismo y distinción entre cogniciones ascendentes y descendentes



**Fuente:** Elaboración propia de los autores (2022).

**Tabla N° 1.** Resumen de las ventajas de una conceptualización más enfocada del antropomorfismo

Adaptación	Criterio de calidad afectado	Problema (P), solución (S) y ventaja (V) para la investigación entre humanos y robots
Distinción de antropomorfismo de sus precursores	Estabilidad	<p>P: Si una definición de antropomorfismo engloba la percepción de que un robot se mueve y se comporta como un humano, solo puede aplicarse razonablemente a robots cuyos movimientos y comportamientos se parezcan a los de los humanos, como NAO (Softbank), pero no a robots que se mueven como máquinas, como Cozmo (Digital Dream Labs), o animales, como el robot foca Paro (Sistema Inteligente). Por ejemplo, Zhang <i>et al.</i> (2008) proponen que el antropomorfismo incluye “manierismos y comportamientos perceptibles”, lo cual es un criterio que difícilmente puede aplicarse a un robot zoomórfico, como Paro.</p> <p>S: La conceptualización propuesta del antropomorfismo se centra en las capacidades mentales y excluye la forma o el movimiento de los robots. Más bien, define explícitamente tales conceptos como precursores.</p> <p>R: Esta conceptualización del antropomorfismo se puede aplicar a diferentes tipos de robots.</p>
Distinción de antropomorfismo de sus consecuencias	Adecuación	<p>P: Si se estudia un robot antropomórfico como Nao (Softbank) y una definición de antropomorfismo incluye tanto la atribución de capacidades mentales humanas como la de valor moral a un robot (por ejemplo, Kahn <i>et al.</i>, 2007), las relaciones causales entre los dos conceptos pueden pasarse por alto. En ese caso, los investigadores pueden no considerar la posibilidad de que la atribución de las capacidades mentales humanas puede causar, más que ser parte de, la atribución de valor moral al robot.</p> <p>S: La conceptualización del antropomorfismo acá desarrollada se centra en las capacidades mentales y no se refiere a la atribución de personalidad y valor moral a los robots. Define explícitamente tales conceptos como consecuencias del antropomorfismo.</p> <p>R: Esta perspectiva estimula un pensamiento más matizado y centrado sobre las relaciones entre el antropomorfismo y sus consecuencias.</p>

**Tabla N° 1.** Resumen de las ventajas de una conceptualización más enfocada del antropomorfismo

Distinción de antropomorfismo de conceptos similares	Adecuación	<p>P: Si una definición de antropomorfismo no se demarca claramente de conceptos similares, como presencia social y animacidad, se pueden pasar por alto diferencias importantes en las respuestas psicológicas a los robots. De manera similar, pueden pasarse por alto los efectos diferenciales del antropomorfismo y conceptos afines. Por ejemplo, el concepto de humanidad percibida de Ho y MacDorman (2010) incluye la distinción entre lo inanimado y lo vivo, lo cual crea una superposición conceptual con la animacidad y dificulta los análisis de los distintos efectos de los robots sociales sobre el antropomorfismo y la animacidad.</p> <p>S: Al explicar el núcleo conceptual del antropomorfismo (y al especificar sus principales dimensiones), se facilita la distinción de este concepto con respecto a otros similares.</p> <p>R: Se pueden estudiar los efectos diferenciales del antropomorfismo y conceptos afines sobre otros resultados.</p>
Especificación de dimensiones concretas de antropomorfismo	Precisión	<p>P: Las definiciones existentes de antropomorfismo generalmente brindan una discusión abstracta del antropomorfismo que, a menudo, solo analiza superficialmente subdimensiones y/o solo listas. Ejemplos de subdimensiones como las expuestas por Bartneck <i>et al.</i> (2009) sostienen que “el antropomorfismo incluye la atribución de forma humana, características humanas y comportamiento humano a los robots” (p. 74). No obstante, sin especificar qué formas, características o comportamientos deben considerarse, y sin proporcionar una justificación teórica de cada aspecto del antropomorfismo, la conceptualización y operacionalización concretas se dejan a discreción de los investigadores y, por lo tanto, se pueden pasar por alto dimensiones cruciales.</p> <p>S: La conceptualización desarrollada define cinco dimensiones principales del antropomorfismo, que reduce el margen de interpretación. Además, se proporciona una justificación teórica para las principales dimensiones del antropomorfismo robótico. Al utilizar la ToM como base, se establece que esta perspectiva del antropomorfismo es integral y refleja esas cinco capacidades mentales que se ha demostrado que son esenciales en la interpretación y predicción del comportamiento de los agentes humanos: es decir, pensar, sentir, percibir.</p> <p>R: Se facilita una conceptualización y operacionalización más consistente y completa del antropomorfismo.</p>

**Fuente:** Elaboración propia de los autores (2022).

En tercer lugar, es posible que sea necesario complementar una medida de antropomorfismo explícito con medidas implícitas en la IHR para obtener una comprensión integral del concepto. Los principales candidatos para evaluar el antropomorfismo implícito son las medidas indirectas de autoinforme (Kim y Sundar, 2012) y las neurológicas (por ejemplo, Urquiza-Haas y Kotrschal, 2015), dada la evolución en la comprensión de los fundamentos neurológicos de la IHR, en general, y del antropomorfismo en particular (Hortensius y Cross, 2018; Urquiza-Haas y Kotrschal, 2015). Se sostiene

que la combinación de autoinforme indirecto y medidas neurológicas ayudará a comprender mejor el antropomorfismo.

En resumen, este estudio ha distinguido el antropomorfismo robótico de sus precursores y consecuencias y ha resaltado sus dimensiones centrales. En el contexto de avances importantes en la robótica social, esperamos que esta conceptualización ayude a aclarar qué es el antropomorfismo y qué no es. En última instancia, se intuye que esta perspectiva más centrada agudizará y sistematizará los estudios en IHR

para contribuir a un conocimiento verdaderamente acumulativo sobre este importante tema.

## Fondos

La investigación presentada en este artículo fue posible gracias a la financiación del Consejo Europeo de Investigación (ERC) en el marco del programa de investigación e innovación Horizonte 2020 de la Unión Europea, en virtud del acuerdo de subvención no. 682733 al segundo autor.

## Notas

Se considera que la conceptualización del antropomorfismo desarrollada en este artículo es una condición suficiente, aunque no necesaria, para la aplicación de la ToM en un robot. La antropomorfización (es decir, la atribución explícita de capacidades mentales humanas a un robot) facilita la aplicación de la ToM a dicho robot. Sin embargo, la aplicación de ToM a un robot no supone que las capacidades mentales humanas se hayan atribuido explícitamente al robot, ya que los procesos implícitos pueden provocar ToM (véase el apartado sobre investigaciones futuras y limitaciones).

## Referencias

Airenti, G. (2018). *The development of anthropomorphism in interaction: Intersubjectivity, imagination, and theory of mind*. *Frontiers in Psychology*, 9, 2136. <https://doi.org/10.3389/fpsyg.2018.02136>

Banks, J. (2020). *Theory of mind in social robots: Replication of five established human tests*. *International Journal of Social Robotics*, 12(2), 403–414. <https://doi.org/10.1007/s12369-019-00588-x>

Barco, A.; de Jong, C.; Peter, J., Kühne, R. y van Straten, C. L. (2020). *Robot morphology and children's perception of social robots. An exploratory study*. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 125–127). <https://doi.org/10.1145/3371382.3378348>

Bartneck, C.; Kuli, D.; Croft, E. y Zoghbi, S. (2009). *Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots*. *International Journal of Social Robotics*, 1(1), 71–81. <https://doi.org/10.1007/s12369-008-0001-3>

Baumeister, R. (2016). *Toward a general theory of motivation: Problems, challenges, opportunities, and the big picture*. *Motivation and Emotion*, 40(1), 1–10. <https://doi.org/10.1007/s11031-015-9521-y>

Beer, J.; Fisk, A. y Rogers, W. (2014). *Toward a framework for levels of robot autonomy in human-robot interaction*. *Journal of Human-Robot Interaction*, 3(2), 74–99. <https://doi.org/10.5898/JHRI.3.2.Beer>

Benninghoff, B., Kulms, P., Hoffmann, L. y Krämer, N. (2013). *Theory of mind in human-robot-communication: Appreciated or not?* <https://doi.org/10.17185/DUEPUBLICO/31357>

Bigman, Y., Waytz, A.; Alterovitz, R. y Gray, K. (2019). *Holding robots responsible: The elements of machine morality*. *Trends in Cognitive Sciences*, 23(5), 365–368. <https://doi.org/10.1016/j.tics.2019.02.008>

Biocca, F.; Harms, C. y Burgoon, J. (2003). *Toward a more robust theory and measure of social presence: Review and suggested criteria*. *Presence: Teleoperators and Virtual Environments*, 12(5), 456–480. <https://doi.org/10.1162/105474603322761270>

Bjorklund, D. y Causey, K. (2018). *Children's thinking. Cognitive development and individual differences*. Sage.

Blanchette, I. y Richards, A. (2010). *The influence of affect on higher level cognition: A review of research on interpretation, judgment, decision making and reasoning*. *Cognition & Emotion*, 24(4), 561–595. <https://doi.org/10.1080/02699930903132496>

Breazeal, C. (2003). *Toward sociable robots*. *Robotics and Autonomous Systems*, 42(3–4), 167–175. [https://doi.org/10.1016/S0921-8890\(02\)00373-1](https://doi.org/10.1016/S0921-8890(02)00373-1)



- Broadbent, E. (2017). *Interactions with robots: The truths we reveal about ourselves*. *Annual Review of Psychology*, 68(1), 627–652. <https://doi.org/10.1146/annurev-psych-010416-043958>
- Coltman, T.; Devinney, T.; Midgley, D. y Venaik, S. (2008). *Formative versus reflective measurement models: Two applications of formative measurement*. *Journal of Business Research*, 61(12), 1250–1262. <https://doi.org/10.1016/j.jbusres.2008.01.013>
- De Jong, C.; Peter, J.; Kühne, R. y Barco, A. (2019). *Children's acceptance of social robots: A narrative review of the research 2000–2017*. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 20(3), 393–425. <https://doi.org/10.1075/is.18071.jon>
- Dennett, D. (1987). *The intentional stance*. MIT Press.
- Dennett, D. (1988). *Conditions of personhood*. En M. F. Goodman (ed.), *What is a person?* (pp. 145–167). Humana Press. [https://doi.org/10.1007/978-1-4612-3950-5\\_7](https://doi.org/10.1007/978-1-4612-3950-5_7)
- Duffy, B. (2003). *Anthropomorphism and the social robot*. *Robotics and Autonomous Systems*, 42(3-4), 177–190. [https://doi.org/10.1016/S0921-8890\(02\)00374-3](https://doi.org/10.1016/S0921-8890(02)00374-3)
- Echterhoff, G.; Bohner, G. y Siebler, F. (2006). *Social Robotics und Mensch-Maschine-Interaktion* [“Social Robotics” and human-machine-interaction]. *Zeitschrift für Sozialpsychologie*, 37(4), 219–231. <https://doi.org/10.1024/0044-3514.37.4.219>
- Epley, N. (2018). *A mind like mine: The exceptionally ordinary underpinnings of anthropomorphism*. *Journal of the Association for Consumer Research*, 3(4), 591–598. <https://doi.org/10.1086/699516>
- Epley, N. y Waytz, A. (2010). Mind perception. En S. T. Fiske, D. T. Gilbert y G. Lindzey (eds.), *The handbook of social psychology* (5° ed., pp. 498–541). Wiley.
- Epley, N.; Waytz, A. y Cacioppo, J. (2007). *On seeing human: A three-factor theory of anthropomorphism*. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Eyssel, F. (2017). *An experimental psychological perspective on social robotics*. *Robotics and Autonomous Systems*, 87, 363–371. <https://doi.org/10.1016/j.robot.2016.08.029>
- Eyssel, F.; Hegel, F.; Horstmann, G. y Wagner, C. (2010). *Anthropomorphic inferences from emotional nonverbal cues: A case study*. En 19th International Symposium in Robot and Human Interactive Communication (pp. 646–651). <https://doi.org/10.1109/ROMAN.2010.5598687>
- Eyssel, F.; Kuchenbrandt, D. y Bobinger, S. (2011). *Effects of anticipated human–robot interaction and predictability of robot behavior on perceptions of anthropomorphism*. En Proceedings of the 6th International Conference on Human–Robot Interaction - IHR '11 (p. 61). <https://doi.org/10.1145/1957656.1957673>
- Ferrari, F.; Paladino, M. y Jetten, J. (2016). *Blurring human-machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness*. *International Journal of Social Robotics*, 8(2), 287–302. <https://doi.org/10.1007/s12369-016-0338-y>
- Fink, J. (2012). *Anthropomorphism and human likeness in the design of robots and human–robot interaction*. En S. S. Ge, O. Khatib, J.-J. Cabibihan, R. Simmons y M.-A. Williams (eds.), *Social robotics. 4° International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012, Proceedings* (pp. 199–208). Springer. [https://doi.org/10.1007/978-3-642-34103-8\\_20](https://doi.org/10.1007/978-3-642-34103-8_20)
- Fong, T. ; Nourbakhsh, I. y Dautenhahn, K. (2003). *A survey of socially interactive robots*. *Robotics and Autonomous Systems*, 42(3–4), 143–166. [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X)
- Fraune, M. (2020). *Our robots, our team: Robot anthropomorphism moderates group effects in human–robot teams*. *Frontiers in Psychology*, 11, 1275. <https://doi.org/10.3389/fpsyg.2020.01275>
- Fraune, M.; Oisted, B.; Sembrowski, C.; Gates, K.; Krupp, M. y Sabanovic, S. (2020). *Effects of robot–human versus robot–robot behavior and entitativity on anthropomorphism*

*and willingness to interact*. *Computers in Human Behavior*, 105, 106220. <https://doi.org/10.1016/j.chb.2019.106220>

Fussell, S.; Kiesler, S.; Setlock, L. y Yew, V. (2008). *How people anthropomorphize robots*. En *Proceedings of the 3° International Conference on Human Robot Interaction - IHR'08* (p. 145). <https://doi.org/10.1145/1349822.1349842>

Gong, L. (2008). *How social is social responses to computers? The function of the degree of anthropomorphism in computer representations*. *Computers in Human Behavior*, 24(4), 1494–1509. <https://doi.org/10.1016/j.chb.2007.05.007>

Gray, H.; Gray, K. y Wegner, D. (2007). *Dimensions of mind perception*. *Science* (New York, N.Y.), 315(5812), 619–619. <https://doi.org/10.1126/science.1134475>

Gunkel, D. (2012). *The machine question. Critical perspectives on AI, robots, and ethics*. MIT Press.

Gunkel, D. (2022). *The symptom of ethics: Rethinking ethics in the face of the machine*. *Human-Machine Communication*, 4, 67–83. <https://doi.org/10.30658/hmc.4.4>

Guzman, A. (2018). *What is human-machine communication, anyway?* En A. L. Guzman (ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 1–28). Peter Lang.

Haslam, N. (2006). *Dehumanization: An integrative review*. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, 10(3), 252–264. [https://doi.org/10.1207/s15327957pspr1003\\_4](https://doi.org/10.1207/s15327957pspr1003_4)

Haslam, N. y Loughnan, S. (2014). *Dehumanization and infrahumanization*. *Annual Review of Psychology*, 65(1), 399–423. <https://doi.org/10.1146/annurev-psych-010213-115045>

Haslam, N.; Loughnan, S.; Kashima, Y. y Bain, P. (2009). *Attributing and denying humanness to others*. *European Review of Social Psychology*, 19(1), 55–85. <https://doi.org/10.1080/10463280801981645>

Hegel, F.; Krach, S.; Kircher, T.; Wrede, B. y Sagerer, G. (2008). *Understanding social robots: A user study on anthropomorphism*. En *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 574–579). <https://doi.org/10.1109/ROMAN.2008.4600728>

Higgins, E. (1990). *Personality, social psychology, and person situation relations: Standards and knowledge activation as a common language*. En L. A. Pervin (ed.), *Handbook of personality: Theory and research*. (pp. 301–338). The Guilford Press.

Himma, K. (2009). *Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent?* *Ethics and Information Technology*, 11(1), 19–29. <https://doi.org/10.1007/s10676-008-9167-5>

Ho, C. y MacDorman, K. (2010). *Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices*. *Computers in Human Behavior*, 26(6), 1508–1518. <https://doi.org/10.1016/j.chb.2010.05.015>

Hortensius, R. y Cross, E. (2018). *From automata to animate beings: The scope and limits of attributing socialness to artificial agents*. *Annals of the New York Academy of Sciences*, 1426(1), 93–110. <https://doi.org/10.1111/nyas.13727>

Hubbard, F. (2011). *“Do androids dream?”: Personhood and intelligent artifacts*. *Temple Law Review*, 83(2), 405–474. <https://ssrn.com/abstract141725983>

Kahn, P.; Ishiguro, H.; Friedman, B.; Kanda, T.; Freier, N.; Severson, R. y Miller, J. (2007). *What is a Human?: Toward psychological benchmarks in the field of human-robot interaction studies*. *Social Behaviour and Communication in Biological and Artificial Systems*, 8(3), 363–390. <https://doi.org/10.1075/is.8.3.04kah>

Kiesler, S. y Goetz, J. (2002). *Mental models of robotic assistants*. En *CHI '02 Extended Abstracts on Human Factors in Computing Systems - CHI '02* (pp. 576–577). <https://doi.org/10.1145/506443.506491>



- Kiesler, S.; Powers, A.; Fussell, S. y Torrey, C. (2008). *Anthropomorphic interactions with a robot and robot-like agent*. *Social Cognition*, 26(2), 169–181. <https://doi.org/10.1521/soco.2008.26.2.169>
- Kim, Y. y Sundar, S. (2012). *Anthropomorphism of computers: Is it mindful or mindless?* *Computers in Human Behavior*, 28(1), 241–250. <https://doi.org/10.1016/j.chb.2011.09.006>
- Krausová, A. y Hazan, H. (2013). *Creating free will in artificial intelligence*. En J. Romportl, P. Ircing, E. Zackova, M. Polak y R. Schuster (eds.), *Beyond AI: Artificial golem intelligence. Proceedings of the International Conference Beyond AI 2013* (pp. 96–109). University of West Bohemia.
- LeDoux, J. (1999). *The emotional brain: The mysterious underpinnings of emotional life*. Phoenix.
- Lee-Won, R.; Joo, Y. y Park, S. (2020). *Media equation*. En J. Bulck (ed.), *The International Encyclopedia of Media Psychology* (1° ed., pp. 1–10). Wiley. <https://doi.org/10.1002/9781119011071.iemp0158>
- Lemaignan, S.; Fink, J.; Dillenbourg, P. y Braboszcz, C. (2014). *The cognitive correlates of anthropomorphism*. En 2014 Human–Robot Interaction Conference, Workshop “HRI: A bridge between Robotics and Neuroscience.”
- Leventhal, H. y Scherer, K. (1987). *The relationship of emotion to cognition: A functional approach to a semantic controversy*. *Cognition & Emotion*, 1(1), 3–28. <https://doi.org/10.1080/02699938708408361>
- Liarokapis, M.; Artemiadis, P. y Kyriakopoulos, K. (2013). *Quantifying anthropomorphism of robot hands*. En 2013 IEEE International Conference on Robotics and Automation (pp. 2041–2046). <https://doi.org/10.1109/ICRA.2013.6630850>
- Lombard, M. y Xu, K. (2021). *Social Responses to Media Technologies in the 21st Century: The Media are Social Actors Paradigm*. *Human-Machine Communication*, 2, 29–55. <https://doi.org/10.30658/hmc.2.2>
- Margolin, D. (2021). *The theory of informative fictions: A character-based approach to false news and other misinformation*. *Communication Theory*, 31(4), 714–736. <https://doi.org/10.1093/ct/qtaa002>
- McLeod, J. y Pan, Z. (2005). *Concept explication and theory construction*. En S. Dunwoody, L. B. Becker, D. M. McLeod y G. M. Kosicki (eds.), *The evolution of key mass communication concepts: Honoring Jack M. McLeod*. (pp. 13–76). Hampton Press.
- Morewedge, C.; Preston, J. y Wegner, D. (2007). *Timescale bias in the attribution of mind*. *Journal of Personality and Social Psychology*, 93(1), 1–11. <https://doi.org/10.1037/0022-3514.93.1.1>
- Nass, C. y Moon, Y. (2000). *Machines and mindlessness: Social responses to computers*. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Ruijten, P. (2018). *Perceptions of human-likeness in human–robot interaction research*. Unpublished manuscript.
- Ryan, R. y Deci, E. (2000). *Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being*. *The American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Salem, M.; Eyssel, F.; Rohlfling, K., Kopp, S. y Joublin, F. (2013). *To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability*. *International Journal of Social Robotics*, 5(3), 313–323. <https://doi.org/10.1007/s12369-013-0196-9>
- Scassellati, B. (2002). *Theory of mind for a humanoid robot*. *Autonomous Robots*, 12(1), 13–24. <https://doi.org/10.1023/A:1013298507114>
- Scholl, B. y Tremoulet, P. (2000). *Perceptual causality and animacy*. *Trends in Cognitive Sciences*, 4(8), 299–309. [https://doi.org/10.1016/S1364-6613\(00\)01506-0](https://doi.org/10.1016/S1364-6613(00)01506-0)
- Schroeder, J. y Epley, N. (2016). *Mistaking minds and machines: How speech affects dehumanization and anthropomorphism*. *Journal of Experimental Psychology. General*, 145(11), 1427–1437. <https://doi.org/10.1037/xge0000214>

- Smith, E. y De Coster, J. (2000). *Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems*. *Personality and Social Psychology Review*, 4(2), 108–131. [https://doi.org/10.1207/S15327957PSPR0402\\_01](https://doi.org/10.1207/S15327957PSPR0402_01)
- Stafford, R.; MacDonald, B.; Jayawardena, C.; Wegner, D. y Broadbent, E. (2014). *Does the robot have a mind? Mind perception and attitudes towards robots predict use of an eldercare robot*. *International Journal of Social Robotics*, 6(1), 17–32. <https://doi.org/10.1007/s12369-013-0186-y>
- Sullins, J. (2006). *When is a robot a moral agent?* *International Review of Information Ethics*, 6, 23–30. doi: <https://doi.org/10.29173/irie136>
- Tan, H., Wang, D. y Sabanovic, S. (2018). *Projecting life onto robots: The effects of cultural factors and design type on multi-Level evaluations of robot anthropomorphism*. En 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (pp. 129–136). <https://doi.org/10.1109/ROMAN.2018.8525584>
- Trovato, G. y Eyssel, F. (2017). *Mind attribution to androids: A comparative study with Italian and Japanese adolescents*. En 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (pp. 561–566). <https://doi.org/10.1109/ROMAN.2017.8172358>
- Urquiza-Haas, E. y Kotrschal, K. (2015). *The mind behind anthropomorphic thinking: Attribution of mental states to other species*. *Animal Behaviour*, 109, 167–176. <https://doi.org/10.1016/j.anbehav.2015.08.011>
- Wang, X. y Krumhuber, E. (2018). *Mind perception of robots varies with their economic versus social function*. *Frontiers in Psychology*, 9, 1230. <https://doi.org/10.3389/fpsyg.2018.01230>
- Waytz, A.; Epley, N. y Cacioppo, J. (2010). *Social cognition unbound: Insights into anthropomorphism and dehumanization*. *Current Directions in Psychological Science*, 19(1), 58–62. <https://doi.org/10.1177/0963721409359302>
- Waytz, A.; Gray, K.; Epley, N. y Wegner, D. M. (2010). *Causes and consequences of mind perception*. *Trends in Cognitive Sciences*, 14(8), 383–388. <https://doi.org/10.1016/j.tics.2010.05.006>
- Waytz, A.; Morewedge, C.; Epley, N.; Monteleone, G.; Gao, J. y Cacioppo, J. (2010). *Making sense by making sentient: Effectance motivation increases anthropomorphism*. *Journal of Personality and Social Psychology*, 99(3), 410–435. <https://doi.org/10.1037/a0020240>
- Wellman, H. (1990). *The child's theory of mind*. MIT Press.
- Wellman, H. y Bartsch, K. (1988). *Young children's reasoning about beliefs*. *Cognition*, 30(3), 239–277. [https://doi.org/10.1016/0010-0277\(88\)90021-2](https://doi.org/10.1016/0010-0277(88)90021-2)
- Westermann, R. (2000). *Wissenschaftstheorie und Experimentalmethodik: Ein Lehrbuch zur Psychologischen Methodenlehre [Philosophy of science and experimental methods: A textbook on psychological methodology]*. Hogrefe.
- Yogeeswaran, K.; Złotowski, J.; Livingstone, M.; Bartneck, C.; Sumioka, H. y Ishiguro, H. (2016). *The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research*. *Journal of Human–Robot Interaction*, 5(2), 29. <https://doi.org/10.5898/JHRI.5.2.Yogeeswaran>
- Zhang, T.; Zhu, B.; Lee, L. y Kaber, D. (2008). *Service robot anthropomorphism and interface design for emotion in human–robot interaction*. En 2008 IEEE International Conference on Automation Science and Engineering (pp. 674–679). <https://doi.org/10.1109/COASE.2008.4626532>
- Złotowski, J.; Proudfoot, D.; Yogeeswaran, K. y Bartneck, C. (2015). *Anthropomorphism: Opportunities and challenges in human–robot interaction*. *International Journal of Social Robotics*, 7(3), 347–360. <https://doi.org/10.1007/s12369-014-0267-6>
- Złotowski, J.; Strasser, E. y Bartneck, C. (2014). *Dimensions of anthropomorphism: From humanness to humanlikeness*. En Proceedings of the 2014 ACM/IEEE International Conference on Human–Robot Interaction - IHR '14 (pp. 66–73). <https://doi.org/10.1145/2559636.2559679>