



FACULTAD DE CIENCIAS HUMANAS Y SOCIALES

**La psicología de la mentira: teorías, motivaciones, efectos.
Dinámica de la mentira y variables de personalidad.**

Autora: Patricia Martín Martínez

Director: Luis De Sebastián Quetglas

Madrid

2025/2026

Índice

Resumen.....	3
Introducción	4
Definiciones	4
Mentira.....	4
Engaño y deshonestidad.....	4
Taxonomía por beneficiario	5
Taxonomía por motivación	6
Modelo teórico.....	7
Modelo ADCAT de Walczyk	7
Teoría de Mantenimiento del Autoconcepto (TMA) de Mazar et ál.	9
Truth-Default Theory (TDT) de Levine	10
Modelo HEXACO: Honestidad-Humildad	10
Conclusión	11
Discusión.....	12
Causas de la conducta mentirosa.....	12
Mentiras auto-protectoras: self-oriented y protectoras.....	13
Mentiras de ganancia personal: self-oriented y promocionales	15
Mentiras prosociales defensivas: other-oriented y protectoras	19
Mentiras de beneficio social: other-oriented y promocionales	24
Personalidades propensas a la mentira: psicopatía y narcisismo	26
La Tríada Oscura	26
Consecuencias de la mentira	29
Nivel interpersonal	29
Nivel intrapersonal	30
Conclusiones	31
Referencias bibliográficas	34
Anexo	39
Metodología	39
Tabla 1	39

Resumen

La mentira constituye un componente habitual de la vida social cuya comprensión exige un marco explicativo que integre procesos cognitivos, motivaciones situacionales y diferencias disposicionales. El presente trabajo analiza la conducta mentirosa a partir de tres ejes complementarios. En primer lugar, se examinan las causas del engaño mediante la clasificación bidimensional de Cantarero et ál. (2018), que organiza cuatro tipos de mentira cotidiana según beneficiario y orientación motivacional, y se explican los mecanismos subyacentes a través de los marcos teóricos de ADCAT, la Teoría de Mantenimiento del Autoconcepto, la Truth-Default Theory y el modelo HEXACO. En segundo lugar, se analizan a fondo los perfiles de psicopatía y narcisismo como personalidades que presentan una elevada propensión a la conducta mentirosa, con causas y comportamientos específicos. En tercer lugar, se analizan las consecuencias del engaño a nivel interpersonal e intrapersonal, mostrando que su impacto sobre la confianza relacional depende de la tipología de la mentira ya sea *self-oriented* o *other-oriented*. Mientras que las consecuencias intrapersonales son homogéneas independientemente del tipo de mentira, mostrando variabilidad en perfiles de personalidad concretos; las consecuencias interpersonales dependen del tipo de mentira realizada, su motivación y el beneficiario. Los hallazgos convergen en una visión de la deshonestidad como fenómeno regulado internamente, motivacionalmente heterogéneo y psicológicamente costoso.

Palabras clave: mentira, engaño, motivaciones, ADCAT, Teoría de Mantenimiento del Autoconcepto, psicopatía, narcisismo.

La psicología de la mentira: teorías, motivaciones, efectos. Dinámica de la mentira y variables de personalidad.

Introducción

Definiciones

La heterogeneidad en el estudio científico de la mentira requiere delimitar unas fronteras conceptuales claras. Sin una definición precisa y operativa de los diferentes términos que se van a utilizar a lo largo del trabajo, resulta imposible distinguir entre conducta mentirosa u otra, así como identificar sus antecedentes psicológicos específicos, consecuencias diferenciales y motivaciones concretas de dicha conducta.

Mentira

Existe un consenso fundamental respecto a la mentira, transversal a la Psicología y la Filosofía: constituye el acto de comunicar una información falsa (Carson, 2010; DePaulo et ál., 1996; Masip Pallejá et ál., 2004). Sin embargo, no existe un consenso claro respecto a si para considerarse mentira una información falsa debe haber intencionalidad de engaño y búsqueda de la creencia errónea para el receptor, es decir, que la información falsa no corresponda a la realidad que conoce el emisor.

Los principales exponentes de la Psicología respecto a la mentira: DePaulo et ál. (1996) y Masip Pallejá et ál. (2004) convergen en que, para denominar a una comunicación falsa, mentira, es necesario que haya intención de engaño. De esta forma, una información falsa comunicada por un error cognitivo, lingüístico, desinformación o de otra índole constituiría lo que se denomina un error honesto, fuera de la categoría de mentira. Esto es especialmente relevante en ámbito legal (Masip Pallejá et ál., 2004). Por otra parte, Carson (2010), exponente de la Filosofía de la mentira, argumenta que no es necesaria intencionalidad de engaño. Debido a la naturaleza de este trabajo, se considerará la versión mayoritaria en el ámbito de la Psicología: la mentira constituye un acto de comunicación errónea en el que el emisor es plenamente consciente de la falsedad de la información y pretende engañar al receptor. Si bien esta definición es altamente consensuada, resulta necesario distinguir la mentira de conceptos similares que frecuentemente se utilizan como sinónimos.

Engaño y deshonestidad

Los términos deshonestidad (*dishonesty*), engaño (*deception*) y mentira (*lying*) se usan frecuentemente de manera intercambiable en el lenguaje coloquial, sin embargo, se deben matizar, ya que presentan particularidades distintivas. El engaño (*deception*) constituye un concepto que engloba cualquier estrategia comunicativa orientada a inducir percepciones erróneas en el receptor, independientemente del medio utilizado o la modalidad de manipulación

informativa (Masip Pallejá et ál., 2004). Además, a diferencia de la mentira, el engaño a menudo requiere éxito en su propósito de instaurar una falsa creencia en el otro para ser considerado como tal (Carson, 2010).

La *Information Manipulation Theory 2* (IMT2) de McCornack et ál. (2014) propone que el engaño implica la manipulación encubierta de información a lo largo de cuatro dimensiones. En primer lugar, la cantidad de información relevante compartida. En segundo lugar, la calidad que involucra la veracidad de la información compartida. En tercer lugar, la forma, es decir, el modo en el que la información se expresa. Finalmente, la relación en cuanto a cómo de relevante es la información compartida. Cuando un emisor viola cualquiera de estas máximas de forma encubierta, se inducen creencias erróneas en el receptor (McCornack et ál., 2014). La mentira, por su parte, se refiere únicamente a la modificación de la segunda máxima (comunicación deliberada de información falsa), estrategia identificada por McCornack et ál. (2014) como falsificación, sin incluir omisiones, ambigüedades no verbales y engaños gestuales. Esta distinción resulta determinante para delimitar el alcance del presente trabajo: se analizan mentiras verbales (falsificaciones activas), quedando fuera del análisis omisiones calculadas y formas no verbales de engaño.

Una segunda distinción relevante coloca la mentira y el engaño en un grupo más grande de comportamientos deshonestos, ya que todos violan normas de integridad. Sin embargo, se diferencian en puntos clave. La deshonestidad (*dishonesty*) abarca cualquier comportamiento que viole expectativas de honestidad e integridad, incluyendo apropiación indebida, incumplimiento de compromisos, plagio intelectual, y engaño comunicativo, sin requerir necesariamente comunicación verbal ni intención de crear creencias específicas falsas en otros (Mazar y Ariely, 2006). Ejemplos de comportamientos deshonestos son: hacer trampa, que implica la violación de reglas para obtener una ventaja injusta; el fraude, que añade dimensiones legales; o el autoengaño, que implica una representación errónea de la realidad ante la propia mente consciente de forma activa pero inconsciente (Butterworth et ál., 2022). Estas distinciones delimitan el objeto de estudio del presente trabajo: se analizan mentiras interpersonales de carácter verbal, quedando fuera otros comportamientos deshonestos que, aunque relacionados con el engaño y la mentira, presentan dinámicas psicológicas, motivaciones y consecuencias diferenciadas que exceden el alcance de este análisis.

Taxonomía por beneficiario

Una primera taxonomía fundamental para comprender las motivaciones del engaño distingue las mentiras según su beneficiario principal: mentiras autocentradas (*self-oriented*), que protegen o benefician al mentiroso, y mentiras centradas en el otro (*other-oriented*), orientadas a

proteger o beneficiar al receptor o terceros. Esta distinción, la estableció por primera vez DePaulo et ál. (1996) en un estudio que analizó más de 1.500 mentiras cotidianas registradas por estudiantes universitarios y adultos durante periodos de una semana.

Las mentiras autocentradas protegen o benefician al mentiroso a través de mecanismos psicológicos de refuerzo negativo (la evasión de vergüenza o regulación emocional) o positivos (preservación de una autopresentación positiva, o regulación emocional); o materialmente, mediante la obtención de ventajas como la evasión de castigos o el logro de conveniencias personales (DePaulo et ál., 1996). Las mentiras centradas en el otro, por su parte, buscan proteger al receptor de malestar emocional, como críticas hirientes o información perturbadora; o beneficiarlo materialmente. Estas últimas, pueden clasificarse en prosociales (benefician a ambas partes) y altruistas puras (solo benefician al otro, incluso a costa del mentiroso) (E. E. Levine y Schweitzer, 2014). Los datos de DePaulo et ál. (1996) muestran una predominancia clara de mentiras autocentradas en ambos grupos (estudiantes universitarios y adultos), con mayor incidencia en el grupo de edad adulta. En el grupo universitario hay un incremento de 19,7 puntos (45,48 vs. 25,74), mientras que en el grupo de adultos el incremento es de 32,2 puntos (56,68 vs. 24,45). Respecto a las mentiras centradas en los otros, se encuentran estables cerca de los 25 puntos en ambos grupos. Cabe destacar una alta variabilidad individual y que no todo queda explicado por el beneficiario de la mentira ya que su suma no alcanza el 100%.

Esta taxonomía por beneficiario constituye el eje organizador de la Discusión del presente trabajo, donde se analizarán en profundidad las causas psicológicas, situacionales y disposicionales que subyacen tanto a las mentiras autocentradas como a las centradas en el otro.

Taxonomía por motivación

Una segunda taxonomía fundamental para comprender las motivaciones del engaño distingue las mentiras basándose en la Teoría del Enfoque Regulatorio. Aplicando este marco, Cantarero et ál. (2018) distinguen dos tipos de mentira según su motivación: mentiras preventivas cuyo foco se encuentra en la prevención de la posición actual; y las mentiras promocionales cuyo foco se encuentra en la promoción y mejora del estatus actual, obteniendo una ganancia adicional que no se obtendría con la verdad.

Las mentiras promocionales responden a lógica de promoción: el emisor identifica una oportunidad de mejora mediante el engaño. Abstenerse de mentir implica mantener el estado previo sin deterioro, pero clausurando una vía de ganancia potencial. Estas mentiras operan tanto en dominio material (obtener empleos, recursos) como psicológico: DePaulo et ál. (1996) documentaron que el engaño cotidiano persigue frecuentemente ventajas en autopresentación y gestión de identidades sociales más que beneficios tangibles directos.

Las mentiras preventivas operan bajo lógica de prevención: la situación se percibe como amenazante y el engaño funciona defensivamente para evitar pérdidas concretas (reputación, relaciones, sanciones). No mentir conlleva riesgo real de deterioro. Esta función protectora resulta congruente con hallazgos de DePaulo et ál. (1996) sobre integración del engaño en dinámicas de regulación interpersonal ordinaria.

Esta taxonomía por motivación constituye el segundo eje organizador de la discusión del presente trabajo, donde se analizarán en profundidad las causas psicológicas, situacionales y disposicionales que subyacen tanto a las mentiras beneficiosas como a las protectoras.

Modelo teórico

Modelo ADCAT de Walczyk

El modelo ADCAT (*Activation-Decision-Construction-Action Theory* o Teoría de Activación-Decisión-Construcción-Acción) es un marco teórico sociocognitivo diseñado para explicar la cognición de las mentiras consideradas graves o de alto riesgo, es decir aquellas cuyo descubrimiento supone una consecuencia relevante para el mentiroso (Walczyk et ál., 2014). Walczyk et ál. (2014) desarrollaron este modelo partiendo del supuesto de que la conducta mentirosa no requiere procesos cognitivos cualitativamente distintos de los habituales, sino que involucra los mismos mecanismos ejecutivos, emocionales, motivacionales y sociales que operan en cualquier otro tipo de respuesta. Dos constructos son transversales a todos sus componentes: (a) la Teoría de la Mente (ToM), que permite al mentiroso inferir los estados mentales del receptor en cada fase del proceso, y (b) los recursos cognitivos disponibles que modulan la efectividad de la conducta mentirosa y cuya cantidad viene determinada por el nivel de motivación para mentir (a mayor motivación, mayor y más sostenida es la asignación de recursos). Además, pueden verse reducidos cuando las emociones elicítadas por la verdad o por la situación de engaño actúan como carga cognitiva externa y reducen la carga cognitiva disponible para la mentira en sí (Walczyk et ál., 2014).

A partir de estos supuestos, proponen cuatro componentes: activación, decisión, construcción y acción. El componente de activación implica el entendimiento de que se está solicitando una verdad a través de la mentalización (ToM) y, después se recupera información episódica, semántica y emocional de la memoria a largo plazo y se transfiere al buffer episódico de la memoria de trabajo. El componente de decisión, considerado el más relevante del modelo y al que se prestará mayor atención por su papel explicativo directo sobre las causas de la mentira, implica la evaluación cuasi-racional o heurística del valor esperado de decir la verdad frente al de mentir. El componente de construcción supone la elaboración de la mentira aplicando el principio de plausibilidad. El mentiroso construye la alternativa falsa a partir de verdades o

recuerdos episódicos alterados, ajustándola a lo que el receptor ya sabe y a lo que es más probable que crea. Con frecuencia se realizan ensayos previos con el propósito de reducir la carga cognitiva durante la entrega. Finalmente, el componente de acción implica llevar a cabo la conducta verbal y física de mentir; dado que la respuesta normativa ante cualquier solicitud de información es la honestidad, el mentiroso debe recurrir al ejecutivo central para inhibir activamente los indicadores de que no está diciendo la verdad. (Walczyk et ál., 2014)

Componente de decisión. Tras la detección del individuo de que se le solicita una verdad, el componente de decisión permite que el individuo evalúe y determine si compartirá la información solicitada de manera precisa o la manipulará de forma engañosa. De esta forma el modelo ADCAT establece que la mentira no es un acto automático ni homogéneo, sino que es un proceso de evaluación dependiente de los recursos cognitivos y temporales disponibles. Se establecen dos rutas de procesamiento: la ruta cuasi-racional que es la principal y la ruta heurística. (Walczyk et ál., 2014)

En la ruta cuasi-racional o central el individuo ha de tener tiempo suficiente, claridad mental y un nivel de estrés no paralizante; de este modo el individuo evalúa intuitivamente las opciones mediante su Valor Esperado (EV) sopesando las probabilidades de éxito, costes y beneficios de cada alternativa. La corteza orbitofrontal asigna valor afectivo a los posibles resultados mientras la memoria de trabajo mantiene activos simultáneamente los objetivos del individuo, el contexto social y la verdad recuperada, permitiendo comparar ambas opciones e inhibir las alternativas incompatibles con esos objetivos o con lo que el receptor ya sabe (ToM). Si el valor esperado de la verdad resulta suficientemente negativo, se activan y evalúan opciones de engaño (moduladas por emociones como el miedo a ser descubierto o el placer anticipado de engañar con éxito) y se elige la alternativa que maximice la utilidad subjetiva. En el caso de que ninguna opción de engaño mejore el EV de la verdad, el individuo optará por la verdad, sin embargo, en caso de que el EV del engaño sea sustancialmente mayor que el EV de la verdad, la motivación para la deshonestidad aumentará y el individuo dispondrá de más recursos cognitivos involucrados en la mentira. (Walczyk et ál., 2014)

La ruta heurística o periférica se activa en caso de que la carga cognitiva disponible sea menor (por ejemplo, alta carga emocional, cansancio o ingesta de sustancias tóxicas) o haya alguna limitación situacional (por ejemplo, tiempo limitado) que impida al proceso ser cuasi-racional. Bajo esta ruta, se recurre a atajos mentales que utilizan reglas preestablecidas como la negación total, imitación de una mentira exitosa del pasado o el uso de la primera mentira suficientemente buena que permita evadir rápidamente el problema. (Walczyk et ál., 2014)

El cálculo cuasi-racional opera directamente sobre las taxonomías definidas en el apartado anterior: la utilidad subjetiva que el individuo asigna a cada resultado determina si la mentira es autocentrada cuando el valor esperado recae sobre beneficios o protecciones propias; o centrada en el otro cuando recae sobre el bienestar del receptor, cuyo daño convierte a la verdad en una opción con valor esperado negativo. A su vez, la naturaleza del desfase entre ambos valores esperados distingue las mentiras preventivas, en las que la verdad presenta un valor esperado marcadamente negativo porque la situación actual se percibe como amenazante, de las mentiras promocionales, en las que la verdad no es necesariamente perjudicial, pero resulta insuficiente para alcanzar una meta superior, de modo que el engaño se elige como vía de ganancia adicional.

Si bien este modelo resulta especialmente operativo para el análisis de situaciones de alto riesgo, la mayoría de las mentiras cotidianas que se analizarán en las secciones siguientes son de baja o media relevancia. Por ello, el presente trabajo no aplica ADCAT como modelo único, sino como marco de referencia del proceso de decisión que se complementará con otros modelos teóricos.

Teoría de Mantenimiento del Autoconcepto (TMA) de Mazar et ál.

La Teoría del Mantenimiento del Autoconcepto de Mazar et ál. (2008) explica por qué personas que se consideran honestas pueden mentir de forma moderada en su vida cotidiana sin que ello modifique negativamente su autoimagen. En cada individuo existe una tensión entre el deseo de obtener algún tipo de ventaja, económica, competitiva o social, y el beneficio interno de preservar un autoconcepto moral positivo; la teoría propone que esta tensión se resuelve mediante una banda de deshonestidad aceptable (o *fudge factor*), un umbral dentro del cual es posible obtener un beneficio sin llegar al punto en que la conducta exija una revisión de la propia identidad moral (Mazar et ál., 2008). Existen dos mecanismos cognitivos que delimitan ese umbral: la maleabilidad de categorización y la desatención de los estándares morales según Mazar et ál. (2008). En primer lugar, la maleabilidad de la categorización implica que el individuo tiende a realizar aquellos actos cuya etiqueta moral resulta lo suficientemente ambigua como para no confrontar su significado deshonesto. Por ejemplo, robar un pintalabios a una amiga resulta categorizable de forma más suave que sustraerle el dinero equivalente, aunque el resultado objetivo sea idéntico. Por otra parte, la desatención a los estándares morales impide que la acción se registre como una violación de la autoimagen, porque en el momento de la conducta deshonesto la persona no activa ni compara su comportamiento con sus propios criterios de honestidad (Mazar et ál., 2008). A estos mecanismos se suma la desconexión moral identificada por Shu et ál. (2009), mediante la cual el individuo reconstruye cognitivamente el acto deshonesto

para neutralizar su carga ética, minimizando el daño causado, difuminando la responsabilidad personal o deshumanizando a la víctima, y el olvido motivado (Shu et ál., 2009), proceso por el que la memoria de las propias transgresiones se suprime activamente con el propósito de preservar una autoimagen coherente con los valores declarados.

En conjunto, estos mecanismos explican el *fudge factor*: cada pequeña falta de honestidad que no altera la forma en que nos vemos a nosotros mismos aumenta un poco lo que consideramos aceptable, creando las bases para un aumento gradual.

Si bien la robustez empírica de estos efectos ha sido cuestionada por estudios recientes que afirman que actualmente los seres humanos están mucho más dispuestos a maximizar su beneficio deshonesto si el contexto les garantiza impunidad social (Guzikevits y Choshen-Hillel, 2022), el constructo sigue siendo el marco más parsimonioso disponible para explicar la deshonestidad moderada de personas generalmente honestas, y es en ese sentido en el que se empleará a lo largo de este trabajo.

Truth-Default Theory (TDT) de Levine

La Truth-Default Theory (TDT) propuesta por T. R. Levine (2014) explica que los seres humanos operan bajo la presunción automática de que la comunicación con los demás es honesta, es decir el estado basal implica asumir pasivamente la veracidad de los mensajes recibidos, a no ser que exista un detonante explícito que active la sospecha. Este estado por defecto debe distinguirse del sesgo de verdad (o *truth-bias*), que se refiere a la tendencia observable de considerar más mensajes como verdaderos que falsos cuando se le pide a alguien que evalúe claramente la veracidad de una comunicación. A diferencia de la asunción de verdad (o *truth-default*), el *truth-bias* opera incluso cuando ya existe cierta sospecha y el individuo sigue inclinándose hacia la creencia de veracidad (T. R. Levine, 2022). El *truth-default* presenta una clara función adaptativa: permite la comunicación eficiente, la transmisión de conocimiento, la cooperación y el mantenimiento de relaciones personales y profesionales sin la carga cognitiva que supondría verificar constantemente la veracidad de cada mensaje recibido (T. R. Levine, 2014). El coste de este estado basal es una vulnerabilidad puntual al engaño, que la TDT considera asumible dado que la mayor parte de la comunicación humana es honesta (Serota y T. R. Levine, 2015). Esta vulnerabilidad es importante para este trabajo por dos razones: se estudiará cómo la psicopatía y el narcisismo utilizan este estado por defecto para conseguir beneficios en sus relaciones con los demás, y se analizará cómo la ruptura del truth-default es la clave para la pérdida de confianza entre personas.

Modelo HEXACO: Honestidad-Humildad

El modelo HEXACO, propuesto por Ashton y Lee (2007), es un marco estructural de la personalidad compuesto por seis dimensiones que surge como evolución del clásico modelo de los Cinco Grandes: Honestidad-Humildad (H), Emocionalidad (E), Extraversión (X), Amabilidad (A), Responsabilidad (C) y Apertura a la experiencia (O). La dimensión de Honestidad-Humildad fue añadida por Ashton y Lee (2007) cuando, mediante varios estudios de lenguaje en diferentes culturas, descubrieron que el modelo de los Cinco Grandes no incluía bien características como la sinceridad, la justicia, la evitación de la codicia y la modestia, que se encuentran en muchas culturas. Mientras los modelos anteriores, ADCAT, TMA y TDT, describen los procesos situacionales mediante los cuales se decide, ejecuta y tolera la mentira, el factor Honestidad-Humildad introduce la dimensión disposicional que explica por qué, ante una misma situación, unos individuos mienten más que otros.

A nivel empírico, la relación entre Honestidad-Humildad y la conducta mentirosa presenta un patrón asimétrico según el tipo de mentira. En el caso de las mentiras autocentradas, la evidencia es consistente: Heck et ál. (2018), a través de un re-análisis de 16 estudios, concluyeron que por cada desviación estándar de aumento en Honestidad-Humildad la razón de posibilidades de hacer trampa se reduce aproximadamente a la mitad ($OR = 0,53$) en situaciones donde mentir reporta un beneficio y la detección es improbable. La relación con las mentiras prosociales, orientadas a beneficiar o proteger a un tercero, presenta más matices. Estudios como el de Paul et ál. (2022) muestran que las personas con mayor puntuación en Honestidad-Humildad tienden a mentir más para no herir los sentimientos de otros; sin embargo, investigaciones posteriores han cuestionado la universalidad de este efecto. Thielmann et ál. (2021) observaron que en tareas de trampa monetaria orientadas a donaciones los individuos altos en Honestidad-Humildad mantienen su comportamiento honesto incluso cuando la mentira beneficiaría a alguien necesitado, y estudios de réplica (Reinhardt et ál., 2024) sugieren que la propensión a decir mentiras prosociales puede estar impulsada por la dimensión de Amabilidad en igual o mayor medida que por la Honestidad-Humildad. Juntos, estos hallazgos muestran que bajos niveles de Honestidad-Humildad predicen de forma robusta la mentira autocentrada, mientras que el comportamiento ante mentiras altruistas depende de la interacción entre Honestidad-Humildad y Amabilidad, así como del contexto específico en que se produce la mentira, distinción que será relevante para analizar los perfiles de psicopatía y narcisismo en la sección de Personalidades Propensas a la Mentira.

Conclusión

Los cuatro modelos presentados conforman un marco explicativo articulado en capas complementarias. ADCAT establece el mecanismo de decisión mediante el cual un individuo

evalúa si merece la pena mentir, integrando variables cognitivas, emocionales y motivacionales en un cálculo cuasi-racional. La Teoría del Mantenimiento del Autoconcepto completa este mecanismo al explicar por qué la mayoría de las personas pueden mentir de forma moderada sin actualizar negativamente su autoimagen, lo que permite dar cuenta de la deshonestidad cotidiana que ADCAT, centrado en situaciones de alto riesgo, no alcanza a cubrir por sí solo. La Truth-Default Theory incorpora la perspectiva del receptor y el contexto social en el que opera la mentira. El estado basal de asunción de honestidad explica por qué la mentira funciona con tanta frecuencia. Finalmente, el factor Honestidad-Humildad del modelo HEXACO añade una dimensión que explica las diferencias personales constantes en el comportamiento mentiroso. Esto servirá como base para comparar la psicopatía y el narcisismo en las secciones siguientes. En conjunto, las taxonomías por beneficiario y por motivación definidas previamente funcionan como sistema de coordenadas sobre el que estos cuatro modelos operan de forma integrada: ADCAT y la TMA explican el proceso y la tolerancia interna de la mentira, la TDT describe las condiciones contextuales que la facilitan, y el factor Honestidad-Humildad determina quién, ante una misma situación, tiene mayor probabilidad de mentir y con qué tipo de motivaciones.

Discusión

Causas de la conducta mentirosa

La afirmación de que la mentira es un componente habitual en la vida social no es una afirmación especulativa: los diarios de conducta muestran que los adultos producen entre una y dos mentiras diarias de media, con una distribución marcadamente asimétrica en la que una minoría de mentirosos prolíficos concentra la mayor parte de los engaños registrados (DePaulo et ál., 1996; Serota y T. R. Levine, 2015). Más allá de los datos autorreportados, la evidencia experimental reciente en Leisge et ál. (2024) demuestra que, ante la oportunidad, cerca de la mitad de las personas eligen el engaño de forma natural a través de tres tareas experimentales diferentes, mostrando una constancia o estabilidad intrapersonal que confirma que la deshonestidad es un rasgo estable de la conducta humana. La pregunta relevante no es, por tanto, si las personas mienten, sino por qué lo hacen, y esa pregunta exige una taxonomía de motivaciones.

Con ese propósito organizador, la presente sección adopta parte de la clasificación bidimensional de Cantarero et ál. (2018), que cruza el eje del beneficiario (*self-oriented* frente a *other-oriented*) con el eje de la orientación motivacional (preventiva frente a promocional), y ordena a partir de esa matriz las nueve motivaciones recogidas en la literatura especializada. El resultado son cuatro cuadrantes que articulan de manera coherente las causas anteriormente

dispersas: mentiras auto-protectoras, mentiras de ganancia personal, mentiras prosociales defensivas y mentiras de beneficio social.

Mentiras auto-protectoras: self-oriented y protectoras

Las mentiras auto-protectoras son aquellas que están orientadas a disminuir un daño que la verdad produciría sobre uno mismo y constituyen la categoría más prevalente en el engaño cotidiano. Los datos de DePaulo et ál. (1996) (obtenidos mediante registros diarios de interacciones) muestran que entre un 26% y un 36% de las mentiras tienen como función principal evitar consecuencias negativas para el propio hablante. Estas pueden ir desde castigos y sanciones explícitas hasta el deterioro de la imagen propia o la exposición de información personal. Por su parte, los datos de Cantarero et ál. (2018) muestran que las mentiras auto-protectoras constituyen el 41% de las mentiras cotidianas diarias. Lo que unifica esta categoría no es el contenido de la mentira sino su lógica subyacente: el individuo percibe que la verdad supone una pérdida, y el engaño opera como mecanismo de evitación. Las dos manifestaciones más habituales de esa lógica son la evitación de castigo y la preservación de la privacidad que comparten el mismo motor cognitivo, pero difieren en el tipo de amenaza que las activa.

Evitación de castigo y consecuencias negativas. La forma más directa de mentira auto-protectora es aquella que responde a una amenaza externa concreta: la posibilidad de ser castigado, reprendido o sancionado. Este subtipo aparece de forma consistente como la motivación más temprana en el desarrollo (las primeras mentiras documentadas en niños tienen casi invariablemente la función de evitación de castigo) y se mantiene a lo largo de la vida adulta en contextos tan distintos como el ámbito laboral, las relaciones de pareja o las interacciones institucionales (DePaulo et ál., 1996; Sai et ál., 2021). La presencia de esta motivación no requiere que el castigo sea severo ni que la detección sea improbable, operando bajo una lógica de prevención de pérdidas: basta con que el individuo evalúe que las consecuencias de decir la verdad superan las de ser descubierto mintiendo.

Lo relevante desde el punto de vista psicológico no es que las personas mientan para evitar castigos, sino que la magnitud y el tipo de mentira que emiten tienden a ajustarse al tamaño de la amenaza percibida (Cantarero et ál., 2018). No se trata de una maximización del engaño sino de una calibración en la que la mentira se construye para ser lo suficientemente creíble como para neutralizar la amenaza, sin exceder lo necesario. Este ajuste sugiere que, incluso en las mentiras más instrumentales, operan restricciones internas que van más allá del cálculo racional de detección (Mazar et ál., 2008; Walczyk et ál., 2014).

Preservación de la privacidad e integridad personal. El segundo subtipo de mentira auto-protectora responde a una amenaza de naturaleza distinta: no la posibilidad de un castigo

externo, sino la exposición de información que el individuo considera parte de su esfera privada o de su identidad (DePaulo et ál., 1996). A diferencia de las mentiras para evitar castigo, que se activan tras una transgresión ya cometida, las mentiras de preservación de la privacidad pueden emitirse en ausencia de cualquier falta previa ya que el hablante miente simplemente porque no quiere revelar algo que considera exclusivamente suyo, como su situación económica, su estado de salud, sus creencias o su vida afectiva (McArthur et ál., 2022).

Esta distinción tiene implicaciones relevantes para la evaluación moral del engaño. Mientras que las mentiras para evitar castigo resultan difíciles de justificar cuando encubren una transgresión real, las mentiras orientadas a proteger la privacidad operan en un espacio ético más ambiguo: la información ocultada no ha sido necesariamente solicitada de forma legítima, y el acto de mentir puede responder a una necesidad genuina de autonomía sobre la propia narrativa identitaria (Carson, 2010). Cantarero et ál. (2018) sitúan esta función dentro de las mentiras protectoras centradas en uno mismo, las cuales gozan de una mayor aceptabilidad social. Esta permisividad se explica porque, como señalan otros autores (Smetana y Wainryb, 2021), en estos casos el receptor cruza un límite al indagar en asuntos personales, y el acto de mentir se comprende como un ejercicio legítimo para proteger la autonomía frente a alguien que exige información que no tiene derecho a conocer (Carson, 2010).

Mecanismo común: el cálculo de amenaza en ADCAT. Pese a diferir en el tipo de amenaza que las activa, ambos subtipos de mentiras auto-protectoras comparten el mismo proceso cognitivo subyacente, descrito por el modelo ADCAT de Walczyk et ál. (2014). Según este marco, ante cualquier pregunta o situación comunicativa, la respuesta verdadera se activa de forma automática e involuntaria en primer lugar desde la memoria a largo plazo: la verdad no requiere elaboración, es la opción por defecto. Como complementa la *Information Manipulation Theory 2* (IMT2) de McCornack et ál. (2014), la decisión de mentir no precede a ese proceso, sino que lo interrumpe: el individuo evalúa, en la fase de decisión, si las consecuencias de emitir esa respuesta verdadera son aceptables o no.

Es en esa evaluación donde reside el mecanismo diferencial de las mentiras auto-protectoras. La amenaza percibida (ya sea un castigo inminente o la exposición no deseada de información privada) vuelve marcadamente negativo el coste subjetivo de la verdad, desplazando la balanza hacia la construcción de una alternativa engañosa. Las variables que modulan esa evaluación incluyen la severidad anticipada de la consecuencia negativa, la probabilidad percibida de ser detectado y la naturaleza de la relación con el receptor: a mayor asimetría de poder entre hablante y oyente, mayor es el peso de la amenaza en el cálculo y, por tanto, mayor la probabilidad de que el individuo opte por mentir (Walczyk et ál., 2014). La fase de

construcción genera entonces la mentira concreta y la fase de acción la ejecuta, incorporando la automonitorización activa de señales que podrían delatar la falsedad. Este ciclo, que en mentirosos habituales puede completarse en fracciones de segundo, explica por qué las mentiras auto-protectoras son tan ubicuas: el sistema que las produce es precisamente el mismo que en condiciones ordinarias garantiza la comunicación honesta.

Mentiras de ganancia personal: self-oriented y promocionales

Las mentiras de ganancia personal constituyen el segundo tipo de engaño más frecuente en la vida cotidiana y se encuentran dirigidas a mejorar una situación propia que no es intrínsecamente negativa. A diferencia de las mentiras auto-protectoras, en las que la verdad se percibe como una amenaza que debe evitarse, la verdad no tiene por qué considerarse aquí como una amenaza o algo perjudicial, pero resulta insuficiente para que el individuo alcance una meta superior que la mentira sí que permite alcanzar. Es decir, la lógica motivacional se invierte: la mentira pasa de ser una estrategia de defensa, seguridad o prevención; a una estrategia de promoción y logro en la que el engaño es una vía de ganancia adicional (Cantarero et ál., 2018). Los datos de DePaulo et ál. (1996) (basados en autorregistros) sitúan este tipo de mentiras en torno a una de cada cinco del total diario (representando un 18,6% de las mentiras en estudiantes y un 19,9% en adultos), cifras que concuerdan con las halladas por Cantarero et ál. (2018) que indican que un 19% de las mentiras diarias son de ganancia personal. La relevancia cualitativa va, sin embargo, más allá de su frecuencia: Cantarero et ál. (2018) muestran cómo el maquiavelismo es un predictor positivo de las mentiras de ganancia personal, y a su vez un predictor negativo de las mentiras auto-protectoras lo que indica que la motivación de conseguir una ventaja es esencial. En otras palabras, una mentira de ganancia personal se emite porque el individuo quiere obtener algo, no porque se vea obligado a protegerse. Esta distinción organiza los tres subtipos que se analizan a continuación: el beneficio material y económico, la obtención de control y poder, y la gestión de la imagen y la búsqueda de aceptación social.

Beneficio material y económico. La forma más habitual de obtener un beneficio personal implica alcanzar un recurso material que no se obtendría con la verdad. Esto implica contar mentiras para manipular a otra persona y así conseguir dinero, recompensas u otras ventajas materiales que no se merecen y no se habrían conseguido en otras circunstancias (McArthur et ál., 2022). El engaño es proactivo y deliberado y funciona como una herramienta de manipulación directa sobre un tercero con el objetivo de que el mentiroso incremente sus propios beneficios. Es esencial esta diferencia entre la obtención de recursos materiales directos y la obtención de ventajas competitivas que no son tangibles y pueden incluir un puesto de trabajo o una beca y que caracterizan al siguiente subtipo.

Cabe recalcar que, cuando las personas se encuentran en una situación en la que tienen la posibilidad de mentir para obtener un beneficio económico sin riesgo de ser descubiertos no maximiza sus beneficios (Mazar et ál., 2008). Es decir, aunque tienen la oportunidad de obtener el máximo beneficio económico sin riesgo de ser descubiertos, la inmensa mayoría elige un nivel de engaño entre el 10% y el 20% del máximo posible. Esto se explica por la TMA y no por otras teorías como el modelo económico racional de coste-beneficio, ya que, al aumentar la recompensa, no aumenta el nivel de engaño (Mazar y Ariely, 2006). Por otra parte, tampoco se explica por la hipótesis de “unas pocas manzanas podridas” (*bad apples*). Suele creerse que el impacto económico de la deshonestidad viene dado principalmente por unas pocas personas que mienten al máximo nivel, sin embargo, el análisis experimental ha demostrado que la pérdida económica se produce por un desplazamiento general en el que la inmensa mayoría de las personas comunes hace trampa solo un poco (para no dañar su autoimagen), y solo un porcentaje ínfimo (0,6% en los experimentos) llega a hacer trampa al nivel máximo posible (Mazar et ál., 2008).

Obtención de control, poder e influencia. A diferencia de las mentiras de beneficio económico que buscan directamente una ganancia financiera o material, las mentiras orientadas a la obtención de control, poder o influencia buscan directamente una ganancia instrumental de naturaleza intangible: recompensas sociales, posiciones de poder, estatus o ventajas competitivas sobre los demás (Walczyk et ál., 2014). La ganancia adicional buscada implica, en este caso, controlar a otros individuos o manipular la información que éstos tienen disponible (Masip Pallejá et ál., 2004). En esa misma línea, DePaulo et ál. (1996) abordan este fenómeno desde la presentación del “yo”: las personas, de forma cotidiana, rara vez mienten por ventajas materiales, sino para obtener recompensas psíquicas como la estima, el afecto y el respeto. De este modo, concluyen que este tipo de mentira se utiliza para crear una imagen del “yo” que en ocasiones es levemente ajustada para adecuarse a las circunstancias, y en otras fabrica un nuevo “yo” falso con el objetivo de obtener poder. Este último tipo de engaño identitario, Walczyk et ál. (2014) lo clasifica como una motivación de mentira seria.

Este subtipo de mentira orientada a la obtención de control, poder e influencia es altamente probable en entornos en los que hay asimetría sobre las propias capacidades que solo resulta verificable a posteriori como en los entornos corporativos y laborales competitivos. Un ejemplo es cuando un directivo o líder roba la idea brillante de un subordinado y la presenta como propia para asegurarse un ascenso y un bono económico (O'Reilly y Doerr, 2020). También es prevalente en negociaciones, donde los individuos con una mentalidad competitiva o motivados por la rivalidad utilizan el engaño informativo para dominar y extraer ventajas de su contraparte

(Gaspar et ál., 2022). Se manifiesta de igual forma en entrevistas de trabajo y situaciones de evaluación o interrogatorio, contextos considerados de alto riesgo en los que el individuo miente deliberadamente para asegurar una posición codiciada, poder o estatus social (Walczyk et ál., 2014).

Los mejores predictores de estas mentiras son los rasgos aversivos de la Tríada Oscura, particularmente el Maquiavelismo y el Narcisismo Grandioso. El Maquiavelismo actúa como predictor porque es un rasgo definido operativamente por el deseo explícito de control y estatus: quienes lo presentan mantienen una visión cínica de la moralidad y consideran que la manipulación y el engaño son estrategias interpersonales válidas para explotar a los demás y ganar dominancia en su entorno (Hilbig et ál., 2025). El Narcisismo Grandioso, por su parte, predice estas mentiras porque los individuos que lo presentan, motivados por un elevado sentido de superioridad, derecho y necesidad de admiración, recurren al engaño para sostener y defender su imagen inflada (O'Reilly y Doerr, 2020; Wright et ál., 2015).

Gestión de la imagen y búsqueda de aceptación. La búsqueda de aceptación social se considera una mentira autocentrada (*self-oriented*) y no centrada en el otro porque se cuenta principalmente para el beneficio del propio mentiroso (DePaulo et ál., 1996; McLeod y Genereux, 2008). Si bien estas mentiras tienen lugar en un contexto interpersonal con el fin de agradar a terceros, su finalidad última es responder a las necesidades del individuo de integrarse, recibir valoraciones positivas y protegerse de la crítica o del rechazo social (Cantarero et ál., 2018). En consecuencia, el objetivo final no es el bienestar ni la protección del receptor, sino la conveniencia y la seguridad psicológica del emisor.

Las mentiras de gestión de la imagen persiguen en todos los casos la misma ganancia: ocupar un lugar favorable en la percepción del otro, ya sea evitando ser juzgado negativamente, proyectando una imagen superior a la real o fingiendo afinidad para obtener aprobación. McArthur et ál. (2022) documentan estas formas reflejando sus distintas lógicas subyacentes: la evitación de la evaluación negativa, en la que el engaño protege al sujeto de la vergüenza o el juicio ajeno permitiéndole presentarse como socialmente competente; y el realzamiento de la autopresentación, orientado a impresionar a los demás negando los propios defectos o construyendo una imagen deliberadamente embellecida. A estas se añade la búsqueda de aprobación mediante el fingimiento de acuerdo (en actitudes, opiniones o preferencias) con el fin de parecer afín al interlocutor y obtener su validación (McLeod y Genereux, 2008). Las tres responden a la misma lógica promocional última: no mantener un *statu quo*, sino alcanzar un estado de aceptación que sin el engaño no estaría garantizado. Cabe señalar que esta lógica puede producir una paradoja: cuando la verdad resulta excesivamente favorable al individuo, el deseo

de proyectar una imagen de equidad o modestia puede llevarlo a reportar resultados inferiores a los reales, de modo que la gestión de la imagen no solo genera exageraciones sino también supresiones estratégicas de logros propios que, de comunicarse, serían percibidos como sospechosamente ventajosos (Guzikevits y Choshen-Hillel, 2022).

A nivel individual, diversos rasgos amplifican de forma específica la probabilidad de recurrir a este subtipo. El primero es el automonitoreo. Los individuos con puntuaciones altas en este rasgo regulan activamente su comportamiento expresivo para adecuarse a las expectativas de cada situación social. Esto los hace especialmente propensos a moldear su autopresentación mediante el engaño cuando la imagen honesta no es suficiente (McLeod y Genereux, 2008). El segundo es el neuroticismo (evaluado en el modelo de los Cinco Grandes). La mayor sensibilidad a la ansiedad, la vergüenza y el castigo social que caracteriza a las personas con puntuaciones altas en este rasgo, incrementa su tendencia a mentir para evitar situaciones de evaluación negativa, encajar en el grupo y reducir la incomodidad social (McArthur et ál., 2022). Ambos rasgos actúan sobre la misma variable: el coste subjetivo percibido de mostrarse tal como se es.

Los tres subtipos descritos: el beneficio material, la obtención de poder y la gestión de la imagen comparten un regulador interno que explica por qué la deshonestidad cotidiana tiende a ser moderada y calibrada antes que maximizada, y cuyo mecanismo se examina a continuación.

Mecanismo común: el *fudge factor* y sus límites. El comportamiento deshonesto en los tres subtipos descritos está regulado por dos circuitos psicológicos que operan en paralelo. El primero es el autoconcepto interno, descrito por la Teoría del Mantenimiento del Autoconcepto de Mazar et ál. (2008) que describe que las personas toleran una banda limitada de deshonestidad que no les obliga a actualizar negativamente su imagen moral interna. El segundo es la imagen social, orientado hacia el exterior y centrado en cómo el individuo es evaluado por los demás (Guzikevits y Choshen-Hillel, 2022). Ambos circuitos suelen operar de forma alineada, pero pueden disociarse: alguien puede mentir no porque necesite preservar su autoconcepto interno, sino únicamente porque necesita gestionar su reputación ante los demás para no parecer deshonesto, o a la inversa. Esta disociación tiene implicaciones relevantes para comprender las diferencias individuales en la conducta mentirosa: en los mentirosos ocasionales, ambos circuitos actúan como frenos simultáneos que mantienen la deshonestidad dentro de límites moderados; en los mentirosos habituales, en cambio, uno o ambos circuitos han perdido su función reguladora, bien porque el autoconcepto ya no se ve amenazado por el engaño, bien porque la imagen social ha dejado de operar como restricción efectiva (Guzikevits y Choshen-Hillel, 2022; Mazar et ál., 2008).

El mecanismo concreto mediante el cual opera el circuito del autoconcepto interno es lo que Mazar et ál. (2008) denominan el *fudge factor* o la “banda de deshonestidad aceptable” dentro de la cual el individuo puede obtener un beneficio (material, posicional o reputacional) sin verse obligado a actualizar negativamente su autoconcepto como persona honesta. Este mecanismo opera a través de dos procesos cognitivos complementarios (Mazar et ál., 2008). El primero es la maleabilidad de la categorización: el sujeto reencuadra la acción en términos compatibles con su autoimagen moral ("es solo una exageración", "todo el mundo lo hace", "no le perjudica a nadie"). De modo que, cuanto más ambigua resulta la etiqueta moral del acto, mayor es la magnitud de deshonestidad que puede tolerarse sin activar una revisión de la propia identidad. El segundo es la desatención a los estándares morales: en el momento de actuar, el individuo simplemente no compara su conducta con sus propios criterios de honestidad, con lo que la acción no llega a registrarse como una violación de la autoimagen. A estos mecanismos, Shu et ál. (2009) añaden una extensión temporal basada en la Teoría de la Desconexión Moral: una vez cometido el acto, el individuo lo reconstruye cognitivamente para neutralizar su carga ética (minimizando el daño causado, difuminando la responsabilidad personal o suprimiendo activamente el recuerdo de la transgresión mediante el olvido motivado), lo que no solo explica la ocurrencia puntual de la deshonestidad sino su persistencia en el tiempo. El límite del mecanismo reside en la propia magnitud del engaño: cuando esta es demasiado grande para ser reencuadrada, la categorización falla y el autoconcepto exige corrección, lo que explica por qué la deshonestidad ordinaria tiende a ser calibrada antes que maximizada. Cada pequeña transgresión que no alcanza ese umbral puede, no obstante, ampliar levemente la banda tolerable, sentando las bases del proceso de escalada gradual o *slippery slope*.

Mentiras prosociales defensivas: other-oriented y protectoras

En las mentiras prosociales defensivas el beneficiario ya no es el mentiroso, es el otro. Se pasa de motivaciones egoístas o autocentradas donde el individuo engaña para protegerse o ganar algo para sí mismo, a motivaciones centradas en el otro (*other-oriented*), en las que la mentira se emite pensando en el bienestar, los intereses o la protección del receptor o de una tercera persona (Cantarero et ál., 2018). De acuerdo con el modelo de Cantarero et ál. (2018) una mentira protectora orientada al otro se define como aquella caracterizada por los intentos del emisor de evitar que un daño, dolor o consecuencia negativa recaiga sobre otra persona. La mentira está impulsada por un enfoque de prevención y aversión a la pérdida; el mentiroso considera que, si dice la verdad, la situación emocional o material de la otra persona empeorará, por lo que miente para mantener la seguridad y el *status quo* del otro.

Esta distinción resulta crucial para delimitar el presente cuadrante del siguiente: en las mentiras prosociales defensivas el impulso es reactivo, existe una amenaza ya configurada para el receptor que la mentira neutraliza; en las mentiras de beneficio social (*other-oriented* y promocional), en cambio, el impulso es prospectivo y no hay daño inminente que evitar sino un resultado positivo nuevo que construir. Las personas no procesan ni juzgan ambos tipos de manera equivalente, percibiendo las mentiras protectoras como significativamente más aceptables y éticas, y relacionándolas positivamente con la empatía, llegando incluso a ser evaluadas como más éticas y morales que la propia verdad objetiva (E. E. Levine y Schweitzer, 2014). En cuanto a su prevalencia, las mentiras protectoras orientadas al otro representaron el 15% del total de mentiras cotidianas en el estudio de Cantarero et ál. (2018), frente al 3% de las beneficiosas, y DePaulo et ál. (1996) muestran que aproximadamente una de cada cuatro mentiras se emite para proteger a otra persona. La investigación reciente mediante autoinformes revela que las mentiras altruistas gozan de una altísima prevalencia en la población general, con una mayoría de individuos admitiendo haber recurrido a ellas recientemente (McArthur et ál., 2022). Las dos manifestaciones más habituales de esta lógica protectora son la protección de los sentimientos ajenos y el mantenimiento de la armonía relacional, que se analizan a continuación.

Protección de sentimientos ajenos. La mentira por protección de sentimientos ajenos se activa habitualmente en situaciones sociales donde revelar la verdad cruda u objetiva causaría dolor emocional, preocupación, vergüenza, pérdida de imagen o estrés innecesario al receptor (DePaulo et ál., 1996). La situación plantea un conflicto directo entre la honestidad y el imperativo social de no causar daño (E. E. Levine y Schweitzer, 2014). E. E. Levine y Schweitzer (2014) definen las mentiras altruistas estrictamente como "declaraciones falsas que son costosas para el mentiroso y que se hacen con la intención de engañar y beneficiar a un objetivo" (p. 108). Esta definición conlleva implicaciones teóricas importantes para el estudio del engaño. En primer lugar, requiere un sacrificio o coste para el emisor, es decir no se produce un bien común (E. E. Levine y Schweitzer, 2015), propio de las mentiras de Pareto que identifica Cantarero et ál. (2018) y que no se analizan en este trabajo. En segundo lugar, no es necesariamente una mentira piadosa o blanca, de hecho, las mentiras altruistas pueden darse en contextos de altas consecuencias y gravedad, como cuando un médico distorsiona un pronóstico clínico para dar consuelo a un paciente en estado terminal (E. E. Levine y Schweitzer, 2015). Por último, desvincula el engaño del egoísmo ya que la propia definición hace que el mentiroso tenga costes asociados a la mentira rompiendo el tradicional vínculo inherente de la mentira con lo inmoral (E. E. Levine y Schweitzer, 2014; E. E. Levine y Schweitzer, 2015).

Sin embargo, estas mentiras además de conllevar un coste para el emisor, ser graves y desvincular la inmoralidad con la mentira, conllevan una paradoja mayor: en contra de la creencia tradicional de que el engaño siempre daña y destruye las relaciones, las mentiras altruistas pueden en realidad aumentar la confianza interpersonal. Cuando el receptor descubre que se le ha mentado con el único propósito de ayudarlo o protegerle, valora el cuidado y la buena intención. Esto incrementa su voluntad de mostrarse vulnerable ante el mentiroso, aumentando drásticamente la confianza basada en la benevolencia. Al mismo tiempo, porque el emisor ha violado las reglas y no ha mostrado compromiso con los principios de honestidad y veracidad, la confianza en su integridad se ve afectada. En resumen, cuando hay un conflicto entre ser honesto y ser amable, el engaño puede generar confianza porque muestra al receptor que la persona ha puesto su bienestar antes que sus propios principios morales, aunque esto pueda dañar la confianza en su integridad. Esto tiene importantes implicaciones en el análisis de las consecuencias de las mentiras. (E. E. Levine y Schweitzer, 2014; E. E. Levine y Schweitzer, 2015)

La motivación de proteger los sentimientos de otra persona puede coexistir y encubrir frecuentemente un componente de beneficio propio o egoísta (*self-serving*). Aunque muchas de estas mentiras prosociales nacen de un deseo genuinamente altruista de apoyar a un compañero o protegerle, un mismo acto de engaño puede ser tanto prosocial como egoísta al mismo tiempo. Por ejemplo, cuando un amigo le da a otro un regalo, pero el regalo no le ha gustado, una mentira prosocial defensiva de los sentimientos ajenos sería asegurar que es muy bonito y que el regalo le encanta. Esto no solo protege los sentimientos del amigo, sino que evita una situación incómoda y una posible discusión, por lo que un mismo acto de engaño puede servir tanto para fines prosociales como para la conveniencia propia (E. E. Levine y Schweitzer, 2014).

El rasgo del modelo HEXACO que mejor predice y muestra una aportación única e importante hacia este tipo de mentira (como dar una evaluación positiva falsa a un ensayo mal escrito) es la Amabilidad (Agreeableness) (Paul et ál., 2022). Esto implica que la motivación principal de estas mentiras se fundamenta en la gentileza y la clemencia que caracterizan a las personas con alta Amabilidad, en contraposición a la dureza y la actitud crítica. Estas personas están dispuestas a sacrificar la honestidad estricta para evitar ser duras con los defectos ajenos y proteger al otro mediante respuestas filtradas (más positivas de lo que deberían ser), lo que a su vez podría funcionar como una estrategia que los hace más agradables entre sus pares (Paul et ál., 2022). Respecto a la predictibilidad de estas mentiras, en correlaciones simples tanto la Honestidad-Humildad (HH) como la Amabilidad muestran una asociación positiva (es decir, personas altas en ambos rasgos son propensas a decir estas mentiras prosociales). Sin embargo,

cuando se realizan análisis de regresión múltiple y se controla la varianza compartida, el estudio original de Paul et ál. (2022) encontró que solo la Amabilidad demostraba tener un valor predictivo único y significativo sobre la mentira prosocial. Cabe destacar que una replicación posterior (Reinhardt et ál., 2024) halló que, bajo este control conjunto, ninguna de las dos dimensiones mantenía un valor predictivo significativo, lo que sugiere que el efecto independiente de la Amabilidad podría ser más débil de lo estimado inicialmente. La diferencia principal radica en que la aversión absoluta a la mentira (la veracidad incondicional) no es el núcleo central del rasgo de Honestidad-Humildad. Ambas dimensiones comparten el hecho de que el individuo es cooperativo y se abstiene de explotar a los demás para beneficio propio, pero los individuos con alta HH sí están dispuestos a mentir si con ello ayudan o benefician a otras personas (Paul et ál., 2022; Reinhardt et ál., 2024).

Mantenimiento de la armonía relacional. A diferencia del subtipo de protección de sentimientos ajenos, cuyo objetivo es preservar el bienestar psicológico y el ego del receptor (evitándole dolor emocional, vergüenza o preocupación), en el mantenimiento de la armonía relacional, lo que se protege es la estabilidad de la relación en sí misma y la fluidez de la interacción social. El engaño se utiliza como una herramienta preventiva para evitar tensiones, conflictos, discusiones exhaustivas o situaciones interpersonales incómodas que surgirían de manera inevitable si se revelara una verdad problemática (E. E. Levine y Schweitzer, 2014; E. E. Levine y Schweitzer, 2015; Whitty y Carville, 2008). A través de sus estudios de diarios sobre la mentira en la vida cotidiana, DePaulo et ál. (1996) documentaron empíricamente que el engaño es una táctica fundamental para el manejo fluido de las interacciones, concluyendo que muchas de las mentiras diarias se dicen precisamente para prevenir la tensión y el conflicto, y minimizar las emociones afectadas y la mala voluntad. En la clasificación creada a partir de estas investigaciones, se identificó este tipo de mentiras como motivadas por razones psicológicas. Se definieron como aquellas que buscan evitar preocupaciones, conflictos u otras situaciones incómodas, y también para hacer las cosas más fáciles o agradables para los demás y mostrarse complaciente (DePaulo et ál., 1996). En el marco de la Teoría ADCAT propuesta por Walczyk et ál. (2014), esta motivación corresponde a la octava categoría de su taxonomía: Afectar las Relaciones Interpersonales / Afiliación. Los autores definen esta categoría como aquella en la que la mentira se cuenta con el fin de mantener, aumentar o controlar el grado de intimidad con el otro. De forma paralela, la misma categoría abarca el extremo opuesto: el individuo puede utilizar el engaño como una barrera para terminar o disminuir activamente la interacción con alguien de quien desea distanciarse.

La interacción social ordinaria exige que los individuos se presenten de una manera aceptable ante los demás, estableciendo un contrato implícito donde las impresiones proyectadas funcionan como expectativas y promesas de carácter moral (Goffman, 1959). Dentro de este marco, la omisión estratégica o la mentira piadosa tienen un lugar funcional y socialmente legitimado; son herramientas comunicativas que se utilizan habitualmente para eludir situaciones embarazosas, evitar la tensión y el conflicto, y minimizar los sentimientos heridos de los interlocutores. Por lo tanto, la armonía relacional se mantiene muchas veces gracias a que este contrato de cortesía nos exige no comunicar de forma cruda todo lo que se piensa. La Teoría de la Resiliencia y la Carga Relacional (TRRL) de Afifi et ál. (2020), aplicada al estudio del engaño por Wnęk y Cantarero (2026) traduce este mismo mecanismo al plano de la salud relacional: lo que Goffman (1959) describe como convención de etiqueta necesaria para que la interacción no derive en conflicto, la TRRL lo conceptualiza como una estrategia de protección de los recursos emocionales del receptor. Salvar las apariencias del otro no es solo un acto de cortesía social sino una forma de incrementar sus reservas emocionales, suavizando el impacto de los comentarios potencialmente hirientes y evitando que la relación acumule una carga de estrés innecesaria (Wnęk y Cantarero, 2026). En este sentido, ambos marcos convergen en una misma conclusión: la supervivencia de los vínculos exige con frecuencia sacrificar la honestidad estricta en favor del cuidado mutuo.

Esta funcionalidad social de la mentira piadosa no es solo una convención cultural sino una condición estructural de la comunicación humana. La Truth-Default Theory de T. R. Levine (2014) sostiene que el estado basal de asunción de verdad es lo que hace posible la comunicación eficiente y cooperación social, dado que la mayor parte de la comunicación humana es honesta; sin este estado por defecto, la carga cognitiva de verificar constantemente cada enunciado haría inviable la interacción ordinaria. No obstante, E. E. Levine y Schweitzer (2015) matizan que la utilidad de estas mentiras tiene efectos divergentes y complejos sobre la confianza cuando el engaño sale a la luz. Lejos de destruir irreparablemente la relación, cuando el receptor detecta la intención protectora del engaño, las mentiras prosociales aumentan la confianza basada en la benevolencia, ya que el receptor percibe una preocupación genuina por su bienestar; sin embargo, simultáneamente dañan la confianza basada en la integridad, al vulnerar la fiabilidad estricta de la palabra del emisor.

A diferencia de la protección de sentimientos, que responde a una amenaza emocional puntual, el mantenimiento de la armonía relacional tiene como objeto el vínculo en su conjunto. Pese a diferir en el objeto que protegen, ambos subtipos comparten el mismo motor psicológico subyacente, cuyo mecanismo se examina a continuación.

Mecanismo común: empatía y evitación del daño. El proceso psicológico que comparten ambos subtipos es la empatía: la capacidad de anticipar el malestar del otro y la motivación de reducirlo antes de que ocurra. Este motor difiere cualitativamente del que impulsa los cuadrantes anteriores porque el regulador emocional no reside en el “yo”, sino fuera de él. En términos disposicionales, esto explica por qué la Amabilidad destaca teóricamente para comprender estas mentiras por encima de la Honestidad-Humildad. Mientras que la Honestidad-Humildad rige principalmente la aversión a explotar o manipular a los demás para beneficio propio (Ashton y Lee, 2007; McArthur et ál., 2022), en el núcleo de la Amabilidad se encuentran la indulgencia y la gentileza, rasgos que llevan al individuo a sacrificar la honestidad estricta para proteger al otro de críticas duras (Paul et ál., 2022). Aunque investigaciones iniciales señalaron a la Amabilidad como el mejor predictor independiente de estas mentiras (Paul et ál., 2022), evidencia reciente matiza que ambas dimensiones comparten una profunda base cooperativa que dificulta la honestidad estricta cuando esta causa daño al prójimo (Reinhardt et ál., 2024). La mentira prosocial defensiva constituye así el punto de máxima tensión entre el cuidado del otro (benevolencia) y la honestidad (justicia) como valores morales en conflicto (E. E. Levine y Schweitzer, 2014).

Mentiras de beneficio social: other-oriented y promocionales

El giro fundamental de las mentiras de beneficio social radica en el cambio de enfoque del emisor: se pasa de una motivación basada en la prevención (evitar un daño o pérdida) a una motivación centrada en la promoción que implica lograr un beneficio o avance (Cantarero et ál., 2018). Según los estudios de registro diario de interacciones realizados por Cantarero et ál. (2018), este tipo de mentira es excepcionalmente rara ya que representa apenas el 3% del total de las mentiras cotidianas, frente al 15% que constituyen las mentiras protectoras orientadas a otros. Esto se explica porque las normas sociales consideran mucho más aceptable e indulgente mentir para proteger a alguien de un daño que mentir para otorgarle un premio o beneficio, lo que conlleva para el mentiroso un coste mucho mayor en caso de que su mentira se descubriese (Cantarero et ál., 2018). Los estudios de E. E. Levine y Schweitzer (2015) documentan el efecto contraintuitivo de que el engaño puede fomentar la confianza: cuando una persona descubre que otra le ha mentado con la intención de beneficiarla, el receptor juzga al mentiroso como una persona altamente benevolente.

Fortalecimiento activo de vínculos. El fortalecimiento activo de vínculos relacionales utiliza el engaño de forma proactiva para construir o profundizar una relación que todavía no ha alcanzado su potencial: el sujeto exagera intereses compartidos, finge mayor compatibilidad emocional o embellece experiencias pasadas con el fin de crear una sensación de vínculo más

estrecho. Cuando el receptor percibe que alguien ha sacrificado la honestidad con la intención de beneficiarle, evalúa al emisor como altamente benevolente, lo que aumenta la confianza basada en la benevolencia y sienta bases sólidas de conexión afectiva (E. E. Levine y Schweitzer, 2015). En este cuadrante cobran especial relevancia dos tipos de mentiras prosociales: las mentiras altruistas, que suponen un coste real para el emisor y se emiten con la única intención de beneficiar al receptor, y las mentiras mutuamente beneficiosas, en las que el engaño favorece simultáneamente a ambas partes, siendo el receptor quien gana la sensación de un vínculo más profundo y el emisor quien gana la relación en sí. En ambos casos, la lógica es promocional y prospectiva, es decir, no se neutraliza un daño existente, sino que se genera activamente una ganancia relacional que la verdad no habría producido.

Este mecanismo opera de forma especialmente eficaz en las fases tempranas del vínculo, cuando la preocupación prioritaria de los individuos es establecer seguridad emocional y la amabilidad pesa más que la honestidad estricta como señal de cuidado. Sin embargo, a medida que la relación madura, esta dinámica se invierte: la honestidad se convierte en la señal de intimidad más potente, y la misma mentira que en el inicio aceleraba la conexión puede percibirse en una relación consolidada como condescendiente o manipuladora si se descubre, erosionando precisamente la confianza basada en la integridad que el vínculo ha ido construyendo (E. E. Levine y Schweitzer, 2015).

Mecanismo común: manejo de impresiones. El manejo de impresiones es una estrategia psicológica y comunicativa mediante la cual los individuos intentan influir en cómo los demás los perciben en los encuentros sociales, sin embargo, en este caso el objetivo de gestionar activamente la imagen no es el propio estatus, sino el vínculo o el bienestar del otro (Whitty y Carville, 2008). Las personas tienen un personaje social y controlan estratégicamente las expresiones que emiten para proyectar una imagen aceptable, creíble y deseable de sí mismas ante su audiencia (Goffman, 1959). En el ámbito del engaño, es una de las principales motivaciones para mentir: los individuos emiten falsedades, no para obtener ganancias materiales, sino para reclamar o salvaguardar identidades, aparentar virtudes o encajar en expectativas sociales (DePaulo et ál., 1996). La Teoría del Mantenimiento del Autoconcepto por Mazar et ál. (2008) postula que las personas viven en constante tensión: desean obtener ventajas mediante el engaño, pero simultáneamente necesitan mantener una visión positiva de sí mismas como personas honestas. La maleabilidad de la categorización, en este caso, permite reencuadrar la mentira como “un acto de generosidad”, “lo que necesitaba escuchar” o “una pequeña exageración que le ayudará”. En el cuadrante prosocial, el autoengaño es especialmente eficaz porque la sociedad en general valora más las mentiras que protegen al otro que las que benefician

al “yo”. El manejo de impresiones prosocial aumenta la confianza basada en benevolencia, pero daña la confianza basada en integridad. La utilidad relacional depende de si la relación prioriza la calidez o la honestidad como valores nucleares (E. E. Levine y Schweitzer, 2015).

Personalidades propensas a la mentira: psicopatía y narcisismo

La Tríada Oscura

La Tríada Oscura (*Dark Triad*) es un constructo que agrupa tres rasgos de personalidad subclínicos y socialmente aversivos: la psicopatía, el narcisismo y el maquiavelismo. Aunque son conceptualmente distintos, estos tres rasgos comparten un núcleo oscuro de características comunes, tales como la hostilidad interpersonal, la manipulación, la frialdad emocional, la explotación de los demás para beneficio propio, un sentido grandioso de autoimportancia y una baja empatía (O’Boyle et ál., 2015). En términos de la estructura general de la personalidad, todos los rasgos de la Tríada Oscura muestran una fuerte correlación negativa con la dimensión de Honestidad-Humildad del modelo HEXACO (Lee y Ashton, 2005).

La psicopatía se define por un conjunto de componentes afectivos y conductuales caracterizados por la frialdad emocional y la desinhibición. Quienes la presentan se caracterizan por un bajo control de los impulsos y una inclinación hacia el comportamiento antisocial, pese a ello pueden mostrar con frecuencia un encanto superficial y habilidades sociales que utilizan estratégicamente para estafar, manipular y aprovecharse de los demás (Kowalski et ál., 2024; Lee y Ashton, 2005; O’Boyle et ál., 2015). El narcisismo, por otra parte, es un rasgo caracterizado fundamentalmente por el engrandecimiento extremo del “yo” y una necesidad voraz de atención y admiración social (O’Boyle et ál., 2015; O’Reilly y Doerr, 2020). Este puede presentarse en forma de grandiosidad o en forma vulnerable (Tommasi et ál., 2025). Por último, el maquiavelismo es un rasgo de la personalidad que se caracteriza por un estilo interpersonal estratégico, manipulador y explotador, combinado con una visión profundamente cínica de la naturaleza humana (Aïn et ál., 2013; Christie y Geis, 1970).

El maquiavelismo no se analiza en profundidad en el presente trabajo por razones sustentadas en la evidencia empírica. En primer lugar, su solapamiento con la psicopatía primaria es tan elevado ($r = 0,66$) (Lee y Ashton, 2005) que muchos investigadores los consideran prácticamente indistinguibles a nivel de constructo. En segundo lugar, la investigación más reciente sobre el Factor Oscuro de la Personalidad (Factor D) muestra que la relación del maquiavelismo con el comportamiento deshonesto está completamente explicada por la varianza compartida con ese núcleo común (Factor D) con la psicopatía, suprimiendo su predicción única e independiente (Hilbig et ál., 2025). Finalmente, Jones y Paulhus (2017) indican que su diferencia conductual respecto a la psicopatía radica fundamentalmente en la orientación

temporal y la estrategia. Mientras la psicopatía impulsa la mentira de forma temeraria y a corto plazo, el maquiavelismo se expresa como un engaño calculado a largo plazo para el manejo de la reputación. En consecuencia, el presente trabajo se centra en la psicopatía y el narcisismo, siendo el narcisismo el único rasgo de la Tríada que mantiene un valor predictivo independiente sobre la deshonestidad más allá del Factor D (Hilbig et ál., 2025), y la psicopatía el prototipo clínico del engaño temerario, conformando así, los dos rasgos con mayor especificidad predictiva y relevancia psicológica diferencial en el contexto de la conducta mentirosa.

Psicopatía. El mecanismo causal que explica la propensión de los psicópatas a mentir radica en su carencia de experiencia emocional genuina, específicamente la ausencia de miedo, empatía y culpa, combinada con una fuerte búsqueda de sensaciones o recompensas (Azizli et ál., 2016). Esta frialdad emocional previene la interferencia afectiva durante el engaño, lo que les permite suprimir o simular expresiones de forma más convincente. Además, según el modelo de Walczyk et ál. (2014) las emociones no actuarían como carga cognitiva externa relevante en la fase de acción, lo que permite a las personas con rasgos psicopáticos mentir sin carga cognitiva externa y disminuyendo la posibilidad de ser descubiertos (Whitty y Carville, 2008). Por otra parte, según el modelo ADCAT de Walczyk et ál. (2014), en el cálculo del valor esperado (EV), las emociones y las diferencias de personalidad influyen directamente en el valor de utilidad subjetiva que se asigna a los resultados. Una persona con rasgos psicopáticos carecerá de frenos morales o emocionales, haciendo que el coste subjetivo de mentir sea prácticamente nulo. Al no deducir la culpa de sus ganancias y al valorar altamente las recompensas inmediatas o la emoción del riesgo, el cálculo EV de la mentira casi siempre resultará netamente positivo y superior al de decir la verdad.

La evidencia demuestra que la psicopatía es un constructo heterogéneo y que sus diferentes componentes tienen impactos distintos sobre la habilidad de engañar, por lo que la psicopatía en su totalidad no está uniformemente relacionada con el éxito en el engaño (Michels et ál., 2020; Porter et ál., 2011). Es decir, la ausencia de experiencia emocional genuina reduce la interferencia afectiva que delataría al mentiroso y, por lo tanto, los rasgos interpersonal-afectivos como la manipulación interpersonal y el afecto frío predicen una menor filtración de expresiones emocionales inconsistentes durante el engaño (Porter et ál., 2011). Sin embargo, los rasgos impulsivo-antisociales operan en sentido contrario, pues la impulsividad actúa como un saboteador activo que incrementa la inconsistencia emocional y, por tanto, la detectabilidad (Porter et ál., 2011). Esta disociación interna cuestiona la imagen clínica y popular del psicópata como mentiroso perfecto: los individuos con alta psicopatía global no logran engañar con mayor éxito a evaluadores ingenuos que la población promedio, y las puntuaciones totales del constructo

no predicen de forma significativa una mayor credibilidad percibida (Michels et ál., 2020; Porter et ál., 2011; Wright et ál., 2015).

La conducta mentirosa del individuo con altos niveles de psicopatía responde a una combinación de frialdad emocional, búsqueda de estimulación y utilitarismo puro. Sus mentiras son fundamentalmente instrumentales, pero con un matiz: mienten para obtener recompensas, ganar admiración o evitar castigos, sin que la posibilidad de causar daño a terceros opere como freno (Baughman et ál., 2014). A ello se suma una dimensión hedónica característica: la susceptibilidad al aburrimiento y la desinhibición propias del rasgo llevan al psicópata a mentir en ocasiones por pura estimulación, con el fin de obtener disfrute personal o, simplemente, comprobar si puede salirse con la suya (Azizli et ál., 2016). El hecho de que no experimenten empatía, culpa o remordimiento, elimina la interferencia emocional que detendría o traicionaría al mentiroso común, lo que explica que su engaño sea ubicuo, tanto en interacciones cotidianas de bajo riesgo como en contextos de alto riesgo. En estos últimos, su habilidad para simular emociones no sentidas puede incluso traducirse en sentencias reducidas o permisos de libertad condicional (Porter et ál., 2011). La planificación, sin embargo, queda subordinada al impulso: a diferencia del maquiavelismo, la mentira psicopática tiende a ser oportunista y orientada a la recompensa inmediata antes que a la estrategia a largo plazo (Jones y Paulhus, 2017).

Narcisismo. El narcisismo predispone a la mentira por sus características centrales: sentido extremo de autoimportancia, un sentido inflado de superioridad, la creencia de tener derechos especiales y una necesidad voraz de atención y admiración social (O'Boyle et ál., 2015). Su mecanismo principal que impulsa la conducta deshonesto es la protección y promoción del ego, a menudo a través del autoengaño y el manejo de impresiones (Wright et ál., 2015). La mentira les sirve de manera instrumental para engrandecer sus logros y obtener el trato preferencial que creen merecer los individuos con narcisismo de tipo grandioso (O'Reilly y Doerr, 2020). Mientras que aquellos con narcisismo vulnerable utilizan la mentira como forma de autoprotección para ocultar la vergüenza, evitar la retroalimentación negativa o las críticas, y proteger su ego frente a cualquier amenaza a su frágil autoimagen (Zvi y Elaad, 2018). Mientras el narcisista miente para proteger y sostener su autoimagen grandiosa, apoyándose en el autoengaño hasta creerse sus propias fabricaciones, el psicópata lo hace de forma fría y consciente, motivado por el utilitarismo o el placer de engañar, con la ventaja de que su ausencia de experiencia afectiva elimina la interferencia emocional que delataría a cualquier otro mentiroso (Jones y Paulhus, 2017; Porter et ál., 2011; Wright et ál., 2015).

Un componente central del mecanismo que impulsa la deshonestidad narcisista es el autoengaño, pues para mantener intacto su sentido de grandiosidad frente a la realidad o las

críticas, los narcisistas distorsionan internamente los hechos (Wright et ál., 2015). En el narcisismo grandioso esta mentira sirve para sostener fantasías de éxito ilimitado, poder y la creencia de que son superiores; en cambio, en el narcisismo vulnerable sirve para eludir la vergüenza, bloquear cualquier retroalimentación negativa y proteger su ego ante lo que perciben como amenazas constantes a su autoimagen (O'Reilly y Doerr, 2020; Tommasi et ál., 2025). Esta dinámica adquiere una dimensión interpersonal relevante cuando se considera en relación con la Truth Default Theory de T. R. Levine (2014) ya que, al creerse genuinamente sus propias fabricaciones, el narcisista elimina la carga cognitiva y suprime las señales no verbales que habitualmente activan la sospecha en el receptor. El narcisista grandioso proyecta una confianza absoluta; el vulnerable, una angustia que parece completamente auténtica. En ambos casos, la teoría sugiere que la ausencia de los desencadenantes necesarios para romper el estado de verdad por defecto hace que el oyente permanezca anclado en la asunción de honestidad. Sin embargo, la evidencia empírica reciente cuestiona que esto convierta al autoengaño en un mecanismo eficaz para sostener el engaño externo: los estudios de Wright et ál. (2015) han demostrado que los individuos con alta tendencia al autoengaño no son mejores mentirosos y resultan ser percibidos como menos creíbles y con menor confianza al mentir. Por tanto, aunque el narcisismo se apoye en el autoengaño como escudo interno, esto no garantiza necesariamente el éxito del engaño en la interacción social.

Consecuencias de la mentira

Nivel interpersonal

El principal daño relacional que produce la mentira es la destrucción de la confianza interpersonal, así como la pérdida de credibilidad y el deterioro general del vínculo (E. E. Levine y Schweitzer, 2014). La evidencia empírica demuestra que el simple acto de engañar conlleva importantes costes relacionales y psicológicos para el propio emisor, incluso cuando el engaño es exitoso y pasa totalmente desapercibido (Elaad, 2022). Cuando una mentira no es detectada erosiona la relación haciendo que las relaciones sean más superficiales y desagradables porque las personas reportan sentirse incómodas y experimentar una reducción en su autoestima y un aumento en su afecto negativo cuando mienten y eso hace que experimenten esas interacciones sociales como menos agradables (DePaulo et ál., 1996; Preuter et ál., 2024). Además, el mentir hace que los emisores muestren menos empatía hacia las emociones de los demás y comiencen a definirse a sí mismos en menor medida en función de sus relaciones personales, lo que aumenta la desconexión emocional y la disminución de la empatía (Kennedy y Kray, 2022). Sin embargo, cuando la mentira sí que es detectada se rompe el estado cognitivo predeterminado de confianza

dando lugar a que la interacción con la otra persona pase de ser natural y cooperativa a ser cognitivamente costosa y llena de desconfianza (T. R. Levine, 2022).

La relación entre mentira y confianza es más compleja de lo que la investigación tradicional ha asumido. Mientras que la mentira egoísta, aquella que beneficia al mentiroso a expensas del receptor, causa un daño significativo y duradero que resulta difícil de reparar incluso con disculpas posteriores (Schweitzer et ál., 2006), la mentira prosocial produce el efecto contrario (E. E. Levine y Schweitzer, 2015). La evidencia empírica demuestra que una persona confía más en quien le dice una mentira prosocial que en quien le dice una verdad egoísta, porque al evaluar el carácter moral se prioriza la benevolencia sobre la honestidad estricta (E. E. Levine y Schweitzer, 2014). Esta distinción modula qué tipo de confianza se ve afectada: la mentira prosocial daña la confianza basada en la integridad, pero aumenta la confianza basada en la benevolencia, lo que revela que la intención subyacente importa más que la veracidad empírica del enunciado (E. E. Levine y Schweitzer, 2015).

El daño interpersonal de la mentira es especialmente severo en relaciones estrechas y a largo plazo, donde el descubrimiento no solo penaliza el acto en sí, sino que pone en duda la totalidad del carácter moral del individuo y destruye la expectativa implícita de honestidad y vulnerabilidad compartida (E. E. Levine y Schweitzer, 2015). La magnitud de ese daño, sin embargo, está determinada por la motivación subyacente: las mentiras egoístas generan respuestas de ira, indignación moral y deseo de represalia, causando un daño a la confianza que raramente se recupera por completo; por su parte, las mentiras vindicativas, orientadas explícitamente a perjudicar al otro, representan el extremo más destructivo (E. E. Levine y Schweitzer, 2015). En el polo opuesto, las mentiras prosociales mitigan el daño relacional al señalar que el mentiroso priorizó el bienestar del otro sobre la honestidad estricta, aunque su uso frecuente en contextos románticos suele asociarse a una menor satisfacción relacional, operando más como mecanismo defensivo ante la incapacidad de gestionar la tensión que como señal de vínculo saludable (Wnęk y Cantarero, 2026).

Nivel intrapersonal

La mentira conlleva costes significativos para el emisor a nivel cognitivo y emocional, incluso cuando pasa totalmente desapercibida y logra su objetivo. La evidencia empírica demuestra de forma consistente que mentir provoca una disminución directa en la autoestima de estado y un aumento del afecto negativo (Preuter et ál., 2024). A nivel intrapsicológico, debido a que el engaño es considerado de manera generalizada como un acto inmoral, ejecutarlo amenaza la visión positiva que el individuo tiene de sí mismo. Estudios longitudinales (diarios) revelan que las personas que mienten en un día determinado experimentan una caída real y

medible en su autoestima de estado en comparación con su nivel del día anterior y con su nivel promedio habitual. De manera contraintuitiva, el tipo de mentira no modula este coste psicológico a nivel emocional. Aunque social e individualmente las mentiras orientadas a los demás (prosociales) se perciben cognitivamente como mucho más aceptables y justificables que las mentiras egoístas o autocentradas, los estudios experimentales han demostrado que ambas disminuyen la autoestima y aumentan el afecto negativo exactamente con la misma intensidad (Preuter et ál., 2024).

Este coste cognitivo y emocional no es universal ni permanente. No es universal porque individuos con ciertos rasgos de la personalidad, como la psicopatía o el maquiavelismo, no experimentan la interferencia emocional, la culpa ni el coste moral típico que frena a un mentiroso promedio (Michels et ál., 2020). Por otro lado, no es permanente porque los seres humanos emplean estrategias psicológicas adaptativas para aliviar la disonancia cognitiva y evitar que el engaño destruya su imagen positiva de sí mismos (Shu et ál., 2009). Entre estas estrategias, la Teoría de Mantenimiento del Autoconcepto describe cómo las personas mienten lo suficiente para obtener un beneficio, pero no tanto como para verse obligadas a considerarse deshonestas, logrando este equilibrio mediante la racionalización de sus actos o la falta de atención consciente a sus propios estándares morales en el momento de la transgresión (Mazar et ál., 2008). Este proceso se complementa con la desconexión moral, mediante la cual el individuo redefine el acto, minimiza su responsabilidad o distorsiona sus consecuencias para salir del engaño con la conciencia intacta (Shu et ál., 2009), y con el olvido motivado, que lleva a borrar selectivamente los estándares éticos tras la conducta deshonesto para blindar la autoimagen frente al arrepentimiento. En su forma más extrema, estos mecanismos convergen en el autoengaño: al creerse genuinamente su propia mentira, el individuo elimina tanto la carga cognitiva de mantener una falsedad como las señales fisiológicas que lo delatarían, y se reserva además la posibilidad de alegar ignorancia si la mentira es descubierta (von Hippel y Trivers, 2011).

Conclusiones

El análisis de las causas de la conducta mentirosa revela que el engaño no es un fenómeno excepcional ni patológico, sino una estrategia adaptativa integrada en la vida social ordinaria. Los registros diarios de conducta demuestran que los adultos producen entre una y dos mentiras de media por día (DePaulo et ál., 1996; Serota y T. R. Levine, 2015). La clasificación bidimensional de Cantarero et ál. (2018), que cruza el eje del beneficiario con el de la orientación motivacional, permite organizar coherentemente esta variabilidad en cuatro cuadrantes con prevalencias diferenciadas: las mentiras auto-protectoras constituyen el tipo más frecuente

(41%), seguidas de las mentiras de ganancia personal (19%), las mentiras prosociales defensivas (15%) y las mentiras de beneficio social (3%). Transversalmente, el análisis demuestra que la deshonestidad cotidiana tiende a ser calibrada antes que maximizada por el mecanismo del *fudge factor* descrito por Mazar et ál. (2008), que establece una banda de tolerancia interna dentro de la cual es posible obtener un beneficio sin exigir una actualización negativa del autoconcepto moral, lo que explica que incluso ante incentivos económicos sin riesgo de detección la mayoría de las personas elija un nivel de engaño notablemente inferior al máximo posible. Este hallazgo desplaza el foco explicativo desde los modelos racionales de coste-beneficio hacia los mecanismos de mantenimiento de la identidad moral y la regulación emocional, que operan de forma diferencial según el tipo de mentira y el perfil disposicional del emisor.

El análisis de las personalidades propensas a la mentira revela que la psicopatía y el narcisismo, pese a compartir una correlación negativa con el factor Honestidad-Humildad del modelo HEXACO (Lee y Ashton, 2005) y a pertenecer ambos al constructo de la Tríada Oscura, presentan mecanismos causales y perfiles conductuales claramente diferenciados. En la psicopatía, la propensión al engaño se origina en la ausencia de experiencia emocional genuina: la inexistencia de miedo, culpa y empatía elimina los frenos afectivos que operan como carga cognitiva externa en el proceso de decisión descrito por ADCAT, haciendo que el valor esperado del engaño resulte casi siempre netamente positivo y sin el coste subjetivo que frena al mentiroso promedio (Walczyk et ál., 2014). Sin embargo, la imagen popular del psicópata como mentiroso perfecto no se sostiene empíricamente ya que la psicopatía es un constructo heterogéneo cuyos componentes operan en sentidos opuestos sobre la eficacia engañosa, de modo que los rasgos interpersonal-afectivos reducen la filtración emocional mientras que los rasgos impulsivo-antisociales la incrementan, y los individuos con alta psicopatía global no logran engañar con mayor éxito que la población promedio en condiciones controladas (Michels et ál., 2020; Porter et ál., 2011). En el narcisismo, el mecanismo central es la protección y promoción del ego a través del autoengaño y el manejo de impresiones (Wright et ál., 2015): el narcisista grandioso miente para sostener fantasías de superioridad e importancia, mientras que el narcisista vulnerable lo hace para protegerse de la vergüenza y bloquear la retroalimentación negativa (O'Reilly y Doerr, 2020; Tommasi et ál., 2025). La dimensión teóricamente más relevante de este perfil reside en que, al creerse genuinamente sus propias fabricaciones, el narcisista, podría en teoría, facilitar la explotación del estado de verdad por defecto al suprimir señales delatorias; sin embargo, la evidencia empírica disponible no confirma que esto se traduzca en una mayor eficacia engañosa en la práctica (T. R. Levine, 2022; Wright et ál., 2015).

En cuanto a las consecuencias de la conducta mentirosa: a nivel interpersonal, el principal coste de la mentira es la erosión de la confianza, que adopta formas cualitativamente distintas según la motivación subyacente: mientras que la mentira egoísta genera respuestas de ira e indignación moral cuyo daño raramente se recupera por completo incluso con disculpas posteriores (Schweitzer et ál., 2006), la mentira prosocial produce el efecto paradójico contrario, incrementando la confianza basada en la benevolencia al tiempo que deteriora la confianza basada en la integridad (E. E. Levine y Schweitzer, 2015). Esta asimetría revela que el juicio moral del receptor prioriza la intención subyacente sobre la veracidad empírica del enunciado. Incluso cuando la mentira no es detectada, el daño relacional existe: la incomodidad experimentada por el emisor reduce la calidad percibida de la interacción, disminuye la empatía hacia el receptor y debilita progresivamente la definición del “yo” en función de las relaciones personales (DePaulo et ál., 1996; Kennedy y Kray, 2022). A nivel intrapersonal, la evidencia empírica demuestra de forma consistente que mentir provoca una disminución directa de la autoestima de estado y un aumento del afecto negativo, con independencia de que la mentira sea egoísta o prosocial (Preuter et ál., 2024). Para amortiguar este coste, los individuos recurren a estrategias adaptativas como la desconexión moral, el olvido motivado y, en su forma más extrema, el autoengaño, que suprime la conciencia de la propia transgresión y elimina con ello tanto la carga cognitiva como las señales que podrían delatarla (Shu et ál., 2009; von Hippel y Trivers, 2011).

La integración de los tres objetivos permite formular una conclusión transversal: la mentira es un fenómeno psicológico cuya comprensión exige un marco explicativo multicapa, en el que los procesos de decisión situacionales de ADCAT, los mecanismos de tolerancia interna de la TMA, las condiciones contextuales de la TDT y las diferencias disposicionales del factor Honestidad-Humildad operan de forma articulada y complementaria sobre el sistema de coordenadas que ofrecen las taxonomías por beneficiario y por motivación. Este nivel de articulación abre varias líneas de investigación relevantes. En primer lugar, los estudios revisados presentan limitaciones de validez ecológica, particularmente los paradigmas experimentales de trampa monetaria, que no reproducen la complejidad motivacional ni las asimetrías de poder propias de las interacciones cotidianas de alto riesgo. Finalmente, el creciente desplazamiento de la interacción social hacia entornos digitales plantea preguntas aún sin respuesta sobre si los principios analizados en este trabajo se mantienen en contextos comunicativos donde la ausencia de señales no verbales altera radicalmente las condiciones en las que la mentira opera.

Referencias bibliográficas

- Afifi, T. D., Harrison, K., Zamanzadeh, N., y Acevedo Callejas, M. (2020). Testing the theory of resilience and relational load in dual career families: Relationship maintenance as stress management. *Journal of Applied Communication Research*, 48(1), 5–25. <https://doi.org/10.1080/00909882.2019.1706097>
- Aïn, S. E. A., Carré, A., Fantini-Hauwel, C., Baudouin, J.-Y., y Besche-Richard, C. (2013). What is the emotional core of the multidimensional Machiavellian personality trait? *Frontiers in Psychology*, 4, 454. <https://doi.org/10.3389/fpsyg.2013.00454>
- Ashton, M. C., y Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, 11(2), 150–166. <https://doi.org/10.1177/1088868306294907>
- Azizli, N., Atkinson, B. E., Baughman, H. M., Chin, K., Vernon, P. A., Harris, E., y Veselka, L. (2016). Lies and crimes: Dark Triad, misconduct, and high-stakes deception. *Personality and Individual Differences*, 89, 34–39. <https://doi.org/10.1016/j.paid.2015.09.034>
- Baughman, H. M., Jonason, P. K., Lyons, M., y Vernon, P. A. (2014). Liar liar pants on fire: Cheater strategies linked to the Dark Triad. *Personality and Individual Differences*, 71, 35–38. <https://doi.org/10.1016/j.paid.2014.07.019>
- Butterworth, J., Trivers, R., y von Hippel, W. (2022). The better to fool you with: Deception and self-deception. *Current Opinion in Psychology*, 47, 101385. <https://doi.org/10.1016/j.copsyc.2022.101385>
- Cantarero, K., van Tilburg, W. A. P., y Szarota, P. (2018). Differentiating everyday lies: A typology of lies based on beneficiary and motivation. *Personality and Individual Differences*, 134, 252–260. <https://doi.org/10.1016/j.paid.2018.05.013>
- Carson, T. L. (2010). *Lying and deception: Theory and practice*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199577415.001.0001>
- Christie, R., y Geis, F. L. (1970). *Studies in Machiavellianism*. Academic Press. <https://doi.org/10.1016/C2013-0-10497-7>
- DePaulo, B. M., Wyer, M. M., Kirkendol, S. E., Kashy, D. A., y Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70(5), 979–995. <https://doi.org/10.1037/0022-3514.70.5.979>
- Elaad, E. (2022). Deceptive behavior: Effects of rational thinking, narcissism, and self-assessed lie- and truth-related abilities. *SAGE Open*, 12(2), 21582440221085012. <https://doi.org/10.1177/21582440221085012>

- Gaspar, J. P., Methasani, R., y Schweitzer, M. E. (2022). Deception in negotiations: Insights and opportunities. *Current Opinion in Psychology*, 47, 101436. <https://doi.org/10.1016/j.copsyc.2022.101436>
- Goffman, E. (1959). *The presentation of self in everyday life*. Doubleday.
- Guzikevits, M., y Choshen-Hillel, S. (2022). The optics of lying: How pursuing an honest social image shapes dishonest behavior. *Current Opinion in Psychology*, 46, 101384. <https://doi.org/10.1016/j.copsyc.2022.101384>
- Heck, D. W., Thielmann, I., Moshagen, M., y Hilbig, B. E. (2018). Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making. *Judgment and Decision Making*, 13(4), 356–371. <https://doi.org/10.1017/S1930297500009232>
- Hilbig, B. E., Thielmann, I., y Heck, D. W. (2025). Filling in the missing pieces: Personality traits (un)related to dishonest behavior. *European Journal of Personality*, 39(5), 732–746. <https://doi.org/10.1177/08902070241293621>
- Jones, D. N., y Paulhus, D. L. (2017). Duplicity among the dark triad: Three faces of deceit. *Journal of Personality and Social Psychology*, 113(2), 329–342. <https://doi.org/10.1037/pspp0000139>
- Kennedy, J. A., y Kray, L. J. (2022). Gender similarities and differences in dishonesty. *Current Opinion in Psychology*, 48, 101461. <https://doi.org/10.1016/j.copsyc.2022.101461>
- Kowalski, C. M., Plouffe, R. A., Daljeet, K. N., Trahair, C., Johnson, L. K., Saklofske, D. H., y Schermer, J. A. (2024). A multi-study investigation assessing the potential redundancy among the dark tetrad using a narrowband trait approach. *Scientific Reports*, 14, 67952. <https://doi.org/10.1038/s41598-024-67952-4>
- Lee, K., y Ashton, M. C. (2005). Psychopathy, Machiavellianism, and narcissism in the five-factor model and the HEXACO model of personality structure. *Personality and Individual Differences*, 38(7), 1571–1582. <https://doi.org/10.1016/j.paid.2004.09.016>
- Leisge, K., Kaczmarek, C., y Schaefer, S. (2024). How often do you cheat? Dispositional influences and intrapersonal stability of dishonest behavior. *Frontiers in Psychology*, 15, 1297058. <https://doi.org/10.3389/fpsyg.2024.1297058>
- Levine, E. E., y Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology*, 53, 107–117. <https://doi.org/10.1016/j.jesp.2014.03.005>
- Levine, E. E., y Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, 126, 88–106. <https://doi.org/10.1016/j.obhdp.2014.10.007>

- Levine, T. R. (2014). Truth-default theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, 33(4), 378–392. <https://doi.org/10.1177/0261927X14535916>
- Levine, T. R. (2022). Truth-default theory and the psychology of lying and deception detection. *Current Opinion in Psychology*, 47, 101380. <https://doi.org/10.1016/j.copsyc.2022.101380>
- Masip Pallejá, J., Garrido Martín, E., y Herrero Alonso, C. (2004). Defining deception. *Anales de Psicología*, 20(1), 147–171. <https://revistas.um.es/analesps/article/view/27631>
- Mazar, N., Amir, O., y Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6), 633–644. <https://doi.org/10.1509/jmkr.45.6.633>
- Mazar, N., y Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing*, 25(1), 117–126. <https://doi.org/10.1509/jppm.25.1.117>
- McArthur, J., Jarvis, R., Bourgeois, C., y Ternes, M. (2022). Lying motivations: Exploring personality correlates of lying and motivations to lie. *Canadian Journal of Behavioural Science / Revue Canadienne des Sciences du Comportement*, 54(4), 335–340. <https://doi.org/10.1037/cbs0000328>
- McCornack, S. A., Morrison, K., Paik, J. E., Wisner, A. M., y Zhu, X. (2014). Information manipulation theory 2: A propositional theory of deceptive discourse production. *Journal of Language and Social Psychology*, 33(4), 348–377. <https://doi.org/10.1177/0261927X14534656>
- McLeod, B. A., y Genereux, R. L. (2008). Predicting the acceptability and likelihood of lying: The interaction of personality with type of lie. *Personality and Individual Differences*, 45(7), 591–596. <https://doi.org/10.1016/j.paid.2008.06.015>
- Michels, M., Molz, G., y Maas genannt BERPohl, F. (2020). The ability to lie and its relations to the dark triad and general intelligence. *Personality and Individual Differences*, 166, 110195. <https://doi.org/10.1016/j.paid.2020.110195>
- O’Boyle, E. H., Forsyth, D. R., Banks, G. C., Story, P. A., y White, C. D. (2015). A meta-analytic test of redundancy and relative importance of the Dark Triad and Five-Factor Model of Personality. *Journal of Personality*, 83(6), 644–664. <https://doi.org/10.1111/jopy.12126>
- O’Reilly, C. A., y Doerr, B. (2020). Conceit and deceit: Lying, cheating, and stealing among grandiose narcissists. *Personality and Individual Differences*, 154, 109627. <https://doi.org/10.1016/j.paid.2019.109627>

- Paul, A. A., Lee, K., y Ashton, M. C. (2022). Who tells prosocial lies? A HEXACO model investigation. *Journal of Research in Personality*, 98, 104232. <https://doi.org/10.1016/j.jrp.2022.104232>
- Porter, S., ten Brinke, L., Baker, A., y Wallace, B. (2011). Would I lie to you? “Leakage” in deceptive facial expressions relates to psychopathy and emotional intelligence. *Personality and Individual Differences*, 51(2), 133–137. <https://doi.org/10.1016/j.paid.2011.03.031>
- Preuter, S., Jaeger, B., y Stel, M. (2024). The costs of lying: Consequences of telling lies on liar’s self-esteem and affect. *British Journal of Social Psychology*, 63(2), 894–908. <https://doi.org/10.1111/bjso.12711>
- Reinhardt, N., Mikesch, M., Hoppe, L., y Reinhard, M.-A. (2024). Close replication of Paul, Lee, and Ashton (2022): Who tells prosocial lies? *Journal of Research in Personality*, 112, 104525. <https://doi.org/10.1016/j.jrp.2024.104525>
- Sai, L., Shang, S., Tay, C., Liu, X., Sheng, T., Fu, G., Ding, X. P., y Lee, K. (2021). Theory of mind, executive function, and lying in children: A meta-analysis. *Developmental Science*, 24(5), e13096. <https://doi.org/10.1111/desc.13096>
- Schweitzer, M. E., Hershey, J. C., y Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes*, 101(1), 1–19. <https://doi.org/10.1016/j.obhdp.2006.05.005>
- Serota, K. B., y Levine, T. R. (2015). A few prolific liars: Variation in the prevalence of lying. *Journal of Language and Social Psychology*, 34(2), 138–157. <https://doi.org/10.1177/0261927X14528804>
- Shu, L. L., Gino, F., y Bazerman, M. H. (2009). *Dishonest deed, clear conscience: Self-preservation through moral disengagement and motivated forgetting* (Working Paper No. 09-078). Harvard Business School. <https://www.hbs.edu/ris/Publication%20Files/09-078.pdf>
- Smetana, J. G., y Wainryb, C. (2021). Adolescents’ and emerging adults’ reminisces about emotions in the context of disclosing, concealing, and lying to parents. *Social Development*, 30(4), 1077–1093. <https://doi.org/10.1111/sode.12522>
- Thielmann, I., Hilbig, B. E., y Böhm, R. (2021). Buying unethical loyalty: A behavioral paradigm and empirical test. *Social Psychological and Personality Science*, 12(3), 363–370. <https://doi.org/10.1177/1948550620905218>

- Tommasi, M., Lauriola, M., y Saggino, A. (2025). Unveiling the dark: Exploring the nomological consistency of the short dark triad and dirty dozen scales. *Europe's Journal of Psychology*, 21(2), 93–108. <https://doi.org/10.5964/ejop.12591>
- von Hippel, W., y Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, 34(1), 1–56. <https://doi.org/10.1017/S0140525X10001354>
- Walczyk, J. J., Harris, L. L., Duck, T. K., y Mulay, D. (2014). A social-cognitive framework for understanding serious lies: Activation-decision-construction-action theory. *New Ideas in Psychology*, 34, 22–36. <https://doi.org/10.1016/j.newideapsych.2014.03.001>
- Whitty, M. T., y Carville, S. E. (2008). Would I lie to you? Self-serving lies and other-oriented lies told across different media. *Computers in Human Behavior*, 24(3), 1021–1031. <https://doi.org/10.1016/j.chb.2007.03.004>
- Wnęk, S., y Cantarero, K. (2026). If I cannot bounce back easily, can I handle the truth? Satisfaction in romantic relationships, resilience and preference towards prosocial lies. *The Journal of Social Psychology*, 1–15. <https://doi.org/10.1080/00224545.2025.2607672>
- Wright, G. R. T., Berry, C. J., Catmur, C., y Bird, G. (2015). Good liars are neither “dark” nor self-deceptive. *PLOS ONE*, 10(6), e0127315. <https://doi.org/10.1371/journal.pone.0127315>
- Zvi, L., y Elaad, E. (2018). Correlates of narcissism, self-reported lies, and self-assessed abilities to tell and detect lies, tell truths, and believe others. *Journal of Investigative Psychology and Offender Profiling*, 15(3), 271–286. <https://doi.org/10.1002/jip.1511>

Anexo

Metodología

El presente trabajo se articula en torno a una revisión narrativa de la literatura científica sobre la conducta mentirosa. La estructura del índice surgió de la lectura en profundidad de un conjunto acotado de textos fundamentales que actuaron como eje conceptual del trabajo y determinaron tanto los ejes temáticos como la organización de los apartados. Entre los que destacan DePaulo et ál. (1996), Cantarero et ál. (2018) y los autores de los modelos teóricos seleccionados (Walczyk et ál., 2014; Mazar et ál., 2008; T. R. Levine, 2022; Ashton y Lee, 2007).

A partir de ese núcleo inicial, la búsqueda bibliográfica se amplió de forma progresiva y orientada por sección, combinando búsquedas por palabras clave en las bases de datos académicas de la Universidad con el rastreo de referencias citadas en los propios textos-eje. Los artículos seleccionados fueron incorporados en función de su relevancia para cada apartado concreto, complementando y matizando el marco conceptual establecido por las obras centrales.

Tabla 1

Taxonomía de las motivaciones de la mentira

Taxonomía por: Motivación		
		Preventivas
		Promocionales
Taxonomía por: Beneficiario	Self-Oriented	Auto-Protectoras: Respuesta a una amenaza real o percibida para evitar un daño, castigo o pérdida propia. • Mentiras por protección propia ("No he sido yo quien ha roto el jarrón") • Mentiras por miedo a las consecuencias ("Ya he entregado el informe") • Mentiras para la preservación de la propia identidad ("No he visto ese mensaje")
	Other-Oriented	De Ganancia Personal: Búsqueda proactiva de ganancias exclusivas para el mentiroso sin amenaza previa que la justifique. • Mentiras por dinero • Mentiras para ejercer control ("Todo el equipo está de acuerdo conmigo") • Mentiras para ejercer poder ("Tengo contactos que pueden hundirte profesionalmente")
	Self-Oriented	Prosociales Defensivas: Evitar que un daño o consecuencia negativa recaiga sobre otra persona. • Mentira por defensa de las relaciones
	Other-Oriented	De Beneficio Social: Asegurar una ganancia o resultado positivo para un tercero sin que medie amenaza que la active. • Mentira de aceptación social ("Tu presentación ha sido la mejor con diferencia")

Nota. Elaboración propia en base a Cantarero et ál. (2018) y DePaulo et ál. (1996).

La tabla reproduce la taxonomía 2×2 propuesta por Cantarero et ál. (2018).

El eje beneficiario indica a quién favorece la mentira: las mentiras *self-oriented* benefician principalmente al propio mentiroso, mientras que las *other-oriented* benefician a un tercero.

El eje motivación se articula a partir de la teoría del enfoque regulatorio: las mentiras preventivas (o protectoras) responden a una lógica de evitación de pérdidas, es decir, se emiten para evitar un daño, castigo o consecuencia negativa ya presente o anticipada; las mentiras promocionales (o de beneficio) responden a una lógica de búsqueda de ganancias, emitiéndose de forma proactiva para obtener ventajas en ausencia de amenaza previa.