Filosofía de las ciencias

Las paradojas de la confirmación

Juan R. Loaiza

Departamento de Filosofía

Universidad Alberto Hurtado

20 de agosto de 2024

Introducción

¿Qué significa que una **hipótesis** esté *bien confirmada*?

- Que sea consistente con los hechos.
- Que los hechos sean como la hipótesis dice que son.
- Que exista *evidencia* en su favor.

¿Qué significa que una hipótesis sea consistente con los hechos?

¿Qué significa que la *evidencia* "hable" en favor de una hipótesis?

¿Qué es la *evidencia*?

"Tener *evidencia*" en favor de una proposición es una expresión sobre nuestro grado de *conocimiento*.

Recordemos:

Una cosa es que una proposición P sea **verdadera**, y otra que **sepamos** que es verdadera.

¿Qué proposiciones sabemos que son verdaderas y cuáles no?

- «Este semestre hay más de 200 estudiantes matriculados en la UAH.»
- «Hay una roca azul orbitando Marte.»
- «En este instante hay 8.171.337.708 de personas en el mundo.»

¿Qué es la *evidencia*?

Tener *evidencia* de una proposición es, entonces, tener algún grado de conocimiento de que el mundo es como la proposición dice que es.

- "Proposición" $\Rightarrow$ "**Hipótesis**"
- "Grado de *conocimiento*" $\Rightarrow$ "Grado de **confirmación**"

*Tener evidencia en favor de una **hipótesis** es que que la **hipótesis** esté **confirmada** en algún grado.*

*Tener evidencia en favor de una **hipótesis** H es que H esté **confirmada** en algún grado.*

Definiendo conjuntos relevantes

Hempel sostiene que el problema de la confirmación importa por varias razones.

Por un lado, queremos definir *sobre qué* es una hipótesis H.

Es importante definir cuál es la **evidencia relevante** para H.

- ¿Qué datos vamos a recoger?
- ¿Qué datos sería importante considerar?

También querríamos definir qué es una **instancia** de H.

- ¿Qué cuenta como una instancia de una ley (e.g., una instancia de "gravedad")?
- ¿Qué no cuenta como una instancia relevante?

Ejemplo

H: Los grupos que se asentaron cerca al mar construían más barcos.

- ¿Qué evidencia *confirmaría* esta hipótesis?

Querríamos inicialmente definir:

- ¿Qué cuenta como un "grupo" (e.g., 3 personas, >5 personas, etc.)?
- ¿Qué cuenta como "estar cerca al mar" (e.g., <200m, <500m, etc.)?
- ¿Qué cuenta como un "barco" (e.g., un bote, un barco grande, etc.)?

Fijando estas preguntas, podríamos (en principio) confirmar H.

Lógica del descubrimiento vs. lógica de la justificación

Además de las razones mencionadas, a Hempel le interesa formular principios para la **justificación** de una hipótesis.

Lógica del descubrimiento

- ¿Qué razones (*buenas o malas*) llevaron a creer en H?

Lógica de la justificación

- ¿Existen *buenas* razones para creer en H?

Recordemos que conocer que P implica *creer* que P, pero también estar *justificado/a* en creer que P.

Resumen

Queremos definir qué significa la expresión:

> *(La evidencia)* E *confirma (o desconfirma) (una hipótesis)* H*.*

Eso nos ayudaría a tener claridad sobre:

- ¿Sobre qué tipo de cosas versa H?
- ¿Qué datos o fenómenos debemos estudiar en relación a H?
- ¿Cuándo estaríamos justificada/os en creer en H?

¿Qué método podría servirnos para explicitar estos criterios? ¡La lógica!

Formalismo básico

En **lógica de predicados** (o lógica de primer orden) hablamos de **proposiciones**.

Proposición: Aquello que puede ser *verdadero* o *falso*. (cfr. enunciado, afirmación)

Simbolizamos las proposiciones distinguiendo **sujetos** (en letras minúsculas) y **predicados** (en letras mayúsculas, aunque Hempel usa palabras).

- «*Juan* está dando *clase*.» $\Rightarrow$ $Clase(j)$ o $C(j)$
- «*María lee* un libro.» $\Rightarrow$ $Lee(m)$ o $L(m)$
- «*Juan* le *regala* un libro a *María*.» $\Rightarrow$ $Regala(j)(m)$ o $R(j)(m)$

Siguiendo a Hempel, usaremos *palabras* en lugar de letras mayúsculas.

Operadores lógicos

Podemos también formar **proposiciones complejas** usando operadores.

- «Juan está dando clase **y** María lee un libro.» $\Rightarrow$ $Clase(j) \wedge Lee(m)$
  - ¡Cuidado! Hempel usa "$\cdot$" en lugar de "$\wedge$": $Clase(j) \cdot Lee(m)$
- «Juan **no** está dando clase.» $\Rightarrow$ $\sim Clase(j)$
- «**Si** Juan da clase, **entonces** María lee un libro.» $\Rightarrow$ $Clase(j) \supset Lee(m)$

¿Cómo formalizaríamos las siguientes proposiciones?

«Camila tiene un gato y un perro.»	$Gato(c) \wedge Perro(c)$
«Si Camila tiene un gato, tiene un perro.»	$Gato(c) \supset Perro(c)$
«Si Camila no tiene un gato, no tiene un perro.»	$\sim Gato(c) \supset \sim Perro(c)$

Cuantificadores

A veces queremos generalizar proposiciones.

Juan está dando clase.	$Clase(j)$
María está dando clase.	$Clase(m)$
Todos están dando clase.	$\forall x(Clase(x))$ $(x)Clase(x)$
Alguien está dando clase.	$\exists x(Clase(x))$ $(Ex)Clase(x)$

Sobre la notación

Hempel usa esta notación por varios motivos:

- Sigue la notación de Russell en *Principia Mathematica*.
- Usa E en lugar de $\exists$ para facilidad de la imprenta.

Cuantificando sobre conjuntos

Cuando queremos decir que un conjunto es miembro de otro, podemos expresar esto con cuantificadores.

> *«Los gatos son animales.»*

Esta proposición dice que “Todo lo que sea gato, será un animal.”

> *«Todo lo (x) que sea gato, será un animal.»*

Esto significa que si encontramos un gato, *cual sea que sea*, será un animal.

$$\forall(x)(Gato(x) \supset Animal(x)) \qquad (x)(Gato(x) \supset Animal(x))$$

Equivalencias lógicas

Finalmente, recordemos que dos fórmulas pueden referir a la misma proposición, y decimos entonces que son **equivalentes**.

«Si Camila estudia derecho, estudia también filosofía.»

$\mathit{Derecho}(c) \supset \mathit{Filosofia}(c)$

Esto implica que «no puede ser que estudie derecho y no estudie filosofía».

$\sim(\mathit{Derecho}(c) \wedge \sim\mathit{Filosofia}(c))$

«No es verdad que Pedro estudie derecho y no filosofía.»

$\sim(\mathit{Derecho}(c) \wedge \sim\mathit{Filosofia}(p))$

Esto implica que «si estudia Derecho, estudia también filosofía.»

$\mathit{Derecho}(c) \supset \mathit{Filosofia}(p)$

Podemos decir entonces que proposiciones de la forma $P \supset Q$ son equivalentes a proposiciones de la forma $\sim(P \wedge \sim Q)$.

Una aproximación inicial

Hempel considera una aproximación inicial: *el criterio de Nicod*.

El criterio de Nicod define primero qué es una **hipótesis**, y luego qué sería que esté **confirmada**.

Hipótesis: Un enunciado de la forma $(x)(P(x) \supset Q(x))$.
«(Para cualquier cosa x) Los x que son P son también Q.»

Ejemplos

- «Los árboles son plantas.» $\Rightarrow (x)(Arbol(x) \supset Planta(x))$
- «Las vacunas son efectivas.» $\Rightarrow (x)(Vacuna(x) \supset Efectiva(x))$
- «Las estrellas tienen alta energía.» $\Rightarrow (x)(Estrella(x) \supset AltaEnergia(x))$

Confirmación según Nicod

Según el criterio de Nicod, una instancia (objeto) a...

Confirma la hipótesis H si satisface

$(x)(P(x) \supset Q(x))$
$P(a) \wedge Q(a)$

Desconfirma la hipótesis H si satisface:

$\sim(x)(P(x) \supset Q(x))$
$P(a) \wedge \sim Q(a)$

Ejemplo

Consideremos la hipótesis:

$$\text{«Las vacunas son efectivas.»} \Rightarrow (x)(Vacuna(x) \supset Efectiva(x))$$

- Confirmada por la vacuna v_1 si $Vacuna(v_1) \wedge Efectiva(v_1)$.
- Desconfirmada por la vacuna v_1 si $Vacuna(v_1) \wedge \sim Efectiva(v_1)$

Primer problema: Hipótesis no universales

Hay hipótesis que no son universales (i.e., no son de la forma $(x)(P(x) \supset Q(x))$).

- «Existe un vida en otras estrellas.»
  - $\exists(x)(Vida(x) \wedge Estrellas(x))$
- «Los seres humanos viven un tiempo limitado.»
  - $(x)(Humano(x) \supset \exists t(Tiempo(t) \wedge ViveHasta(x, t)))$
- «Puedes engañar a todos en algún momento y a algunas personas todo el tiempo, pero no puedes engañar a todas las personas todo el tiempo.»
  - $(x)\exists tE(x, t) \wedge \exists x(t)E(x, t) \wedge \sim(x)(t)E(x, t)$

La paradoja de la confirmación

El problema más importante, sin embargo, es que el criterio introduce una **paradoja**.

(S_1) $(x)(\mathit{Cuervo}(x) \supset \mathit{Negro}(x))$
Todos los cuervos son negros.

(S_2) $(x)({\sim}\mathit{Negro}(x) \supset {\sim}\mathit{Cuervo}(x))$
Todo lo no-negro es un no-cuervo.

Es una verdad lógica que $P \supset Q$ es **lógicamente equivalente** a ${\sim}Q \supset {\sim}P$.

Por lo tanto, ¡S_1 **es lógicamente equivalente** a S_2!

¿Y eso por qué es un problema? Veamos.

La condición de equivalencia

Si dos hipótesis son **equivalentes**, sus formulaciones refieren a la misma **proposición** (i.e., al mismo **estado de cosas**).

- Decir "Todo lo no-negro es un no-cuervo" es una forma equivalente de referir a la proposición «Todos los cuervos son negros.»

Y si dos oraciones refieren a la **misma proposición**, la **evidencia** para creer en una debería ser evidencia para creer en la otra.

- La evidencia en favor de «Todos los cuervos son negros» sería evidencia a favor de «Todo lo no-negro es un no-cuervo» y **viceversa**.

Esto introduce resultados contraintuitivos.

¿Qué evidencia confirmaría estas hipótesis?

Todos los cuervos son negros

$(x)(Cuervo(x) \supset Negro(x))$

- $Cuervo(a) \wedge Negro(a)$
- $Cuervo(b) \wedge Negro(b)$
- $Cuervo(c) \wedge Negro(c)$

Todo lo no-negro es un no-cuervo

$(x)(\sim Negro(x) \supset \sim Cuervo(x))$

- $\sim Negro(a) \wedge \sim Cuervo(a))$
- $\sim Negro(b) \wedge \sim Cuervo(b))$
- $\sim Negro(c) \wedge \sim Cuervo(c))$

¡Cualquier objeto que *no sea negro* y *no sea un cuervo* confirma que *todos los cuervos son negros*!

Ejemplo: Científico maligno #1

Supongamos que un científico maligno nos quiere convencer de que:

Todos los perros son animales marinos. $(x)(Perro(x) \supset Marino(x))$

Para ello, el científico recopila la siguiente evidencia:

Una silla a no es marina.	$\sim Marino(a) \wedge \sim Perro(a)$
Un martillo b no es marino.	$\sim Marino(b) \wedge \sim Perro(b)$
Un carro c no es marino.	$\sim Marino(c) \wedge \sim Perro(c)$

Dado que $(x)(Perro(x) \supset Marino(x))$ es equivalente a $(x)(\sim Marino(x) \supset \sim Perro(x))$, ¡toda la evidencia que recogió confirma su hipótesis!

Resumen

Nicod proponía un criterio simple e intuitivo de confirmación:

Criterio N. Una hipótesis H tiene la forma $(x)(P(x) \supset Q(x))$ y es confirmada por todos los objetos $a, b, c, \ldots n$ que $P(a) \wedge Q(a)$.

Sin embargo, esto lleva a una paradoja:

- Todos los objetos $a, b, c, \ldots, n$ que satisfagan $\sim Q(a) \wedge \sim P(a)$ también confirman H.
  - Ej.: Una vaca amarilla confirma la hipótesis «Todos los cuervos son negros.»

¿Qué salió mal? ¿Qué suposición hicimos que no sea la correcta?

Revisitando presuposiciones

Siempre que se llega a una conclusión que consideramos absurda, debemos hacernos dos preguntas.

- ¿Es **falsa** alguna suposición que hice en el camino?
- ¿Es *realmente* **absurda** la conclusión a la que llegué?

Hempel comienza considerando las suposiciones que introdujo en el camino:

1. Una hipótesis científica general puede representarse con la forma $(x)(P(x) \supset Q(x))$.
2. Si H_1 y H_2 son lógicamente equivalentes, la evidencia en favor de H_1 es también evidencia que confirma H_2.

¿Podemos representar hipótesis con condicionales?

Una alternativa que Hempel considera es que la formalización de una hipótesis no sea tan sencilla.

Quizás, por ejemplo, una hipótesis científica tiene oculto algún enunciado existencial:

Todos los gatos son animales. $(x)(G(x) \supset A(x)) \wedge \exists x G(x)$

Sin embargo, no toda hipótesis científica declara la existencia de los objetos a los que refiere.

Una luna con mayor masa que su planeta lo aplastaría.
$(x)(y)(L(x) \wedge P(y) \wedge M(x)(y) \supset A(x)(y))$

Lógica del descubrimiento vs. lógica de la justificación 2.0

Hempel sostiene que la paradoja desaparece si recordamos que una cosa es la evidencia *parezca* confirmar una hipótesis y otra es que *lógicamente* lo haga.

Lógica del descubrimiento

- ¿Qué razones (*buenas o malas*) llevaron a creer en H?

Lógica de la justificación

- ¿Existen *buenas* razones para creer en H?

Propuesta: Podemos aceptar que evidencia extraña confirme una hipótesis, aunque psicológicamente nos parezca contraintuitivo.

¿Por qué la dificultad es psicológica?

Hempel nos pide considerar la siguiente hipótesis:

H_1: *Todas las sales de sodio queman amarillo.* $(x)(S(x) \supset A(x))$

Según la condición de equivalencia y el criterio de Nicod, esto es equivalente a:

H_2: *Todo lo que no quema amarillo no es una sal de sodio.*
$(x)({\sim}A(x) \supset {\sim}S(x))$

Puesta así, la segunda hipótesis no parece tan extraña:

- Quemar algo y revisar si quema amarillo es un test de si es una sal de sodio.
- La evidencia en favor de H_2 parece confirmar (tranquilamente) H_1.

Comparando hipótesis

¿Cuál es la **diferencia** entre estas dos hipótesis entonces?

H_s: Todas las sales de sodio queman amarillo.
$(x)(S(x) \supset (A(x))$

H_c: Todos los cuervos son negros.
$(x)(C(x) \supset N(x))$

Nos parece paradójico el caso de H_c porque creemos *obvio* cómo se identifican los cuervos.

- Es claro que una vaca no es un cuervo. No solicitamos un test ulterior.
- No es tan claro que un material que parezca una sal sea o no una sal de sodio. Hacemos el test de quemarla.

Conocimiento de trasfondo

El punto central de Hempel, entonces, puede resumirse así:

- Lo que **confirma** o **desconfirma** una hipótesis es un asunto ***lógico***, aunque parezca contraintuitivo.
- Lo que nos haga **plausible** una confirmación es un asunto ***psicológico***, y viene acompañado de **conocimiento de trasfondo**.

La ciencia nunca ocurre aislada de nuestras creencias. Esto sugiere dos rutas de análisis:

- **Lógica de la ciencia**: Relaciones lógicas entre enunciados científicos
- **Psicología de la ciencia**: Factores mentales y sociales de la aceptación de hipótesis

Hacia una lógica de la confirmación

Para Hempel, la **confirmación** de una hipótesis comprende tres elementos:

Hipótesis

Dicta sobre qué objetos habla.

Evidencia

Instancias relevantes para la hipótesis.

Conocimiento de trasfondo

Determina la aceptabilidad de la hipótesis.

Lección: Es vital distinguir aspectos **lógicos** y **psicológicos** de la confirmación de una hipótesis.

- Una teoría puede estar bien confirmada aunque nadie crea en ella.

¿Y nos quedamos con la solución de Hempel?

Existen otras soluciones alternativas al problema de la confirmación.

1. Rechazar que evidencia en favor de $(x)(\sim Q(x) \supset \sim P(x))$ confirme $(x)(P(x) \supset Q(x))$
   - Imponer restricciones sobre la referencia (Quine).
   - Distinguir nociones de confirmación *condicional* e *incondicional* (Bayesianismo)
2. Rechazar la distinción entre lógica y psicología de la ciencia.
   - Colapsar *aceptación* con *justificación* (Kuhn, Feyerabend).
   - Rechazar la necesidad de *justificación* al hablar de confirmación (Sociología)

La paradoja de la confirmación

Nos enfocamos en resolver la pregunta:

> *¿Qué significa que una **hipótesis** esté **bien confirmada**?*

Exploramos una respuesta **intuitiva**:

- Una hipótesis científica H tiene la forma $(x)(P(x) \supset Q(x))$.
- H está bien confirmada cuando tenemos **evidencia** de objetos $a, b, c...n$ tales que $P(a) \wedge Q(a)$.

Esto llevó al resultado (aparentemente) **paradójico**:

- H es confirmada por casos $a, b, c...n$ tales que $\sim Q(a) \wedge \sim P(a)$.

Disolviendo la paradoja de la confirmación

La paradoja se disuelve cuando distinguimos:

- ¿Qué **justifica (lógicamente)** aceptar una hipótesis?
- ¿Qué hace **psicológicamente plausible** aceptar una hipótesis?

Podemos aceptar que casos $a, b, c...n$ tales que $\sim Q(a) \land \sim P(a)$ confirmen $(x)(P(x) \supset Q(x))$, aunque **no nos parezca** intuitivo.

La plausibilidad psicológica depende de **conocimiento de trasfondo** sobre el que evaluamos la evidencia.

- La evidencia no se evalúa **aislada** de nuestras creencias.

En el próximo episodio…

Los problemas de la **confirmación** no terminan allí.

- ¿**Cuánta evidencia** es necesaria para justificar una hipótesis?

Esto nos lleva a preguntarnos:

> *¿Cuándo está **justificada** una **inferencia inductiva**?*

Leeremos el análisis de **Goodman** sobre el **problema de la inducción** y veremos cómo atacar problemas de inferencias inductivas (e.g., predicciones).