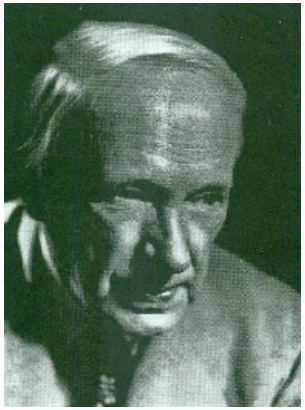
The Moore Sentence and The Fitch Paradox

Hans van Ditmarsch
LORIA / CNRS, Nancy & IMSc, Chennai

hans.van-ditmarsch@loria.fr

Musée de l'Ecole de Nancy

The Moore-sentence $p \wedge \neg Kp$



G.E. Moore. *A reply to my critics.* In P.A. Schilpp, editor, The Philosophy of G.E. Moore, pages 535–677. Northwestern University, Evanston IL, 1942. The Library of Living Philosophers (volume 4).

> *" 'I went to the pictures last Tuesday, but I don't believe that I did' is a perfectly absurd thing to say, although* what *is asserted is something which is perfectly possible logically" (page 543).*

The absurdity follows from the implicature 'asserting φ implies $B\varphi$' pointed out in Moore's Ethics, 1912.

The Moore-sentence $p \wedge \neg Kp$

> " *'I went to the pictures last Tuesday, but I don't believe that I did' is a perfectly absurd thing to say, although* what *is asserted is something which is perfectly possible logically*".

The absurdity follows from implicature 'asserting φ implies $B\varphi$'.

- Proposition p stands for 'I went to the pictures last Tuesday'.
- Write K for the epistemic modal operator, not B.
- 'I went to the pictures last Tuesday, but I don't believe that I did' is formalized as $p \wedge \neg Kp$.
- Absurdity follows from implicature 'asserting φ implies $K\varphi$': $K(p \wedge \neg Kp)$ is inconsistent (for 'usual' knowledge and belief)

The Moore-sentence $p \wedge \neg Kp$

$K(p \wedge \neg Kp)$ is inconsistent.

$K(p \wedge \neg Kp)$
$\Rightarrow$
$Kp \wedge K\neg Kp$
$\Rightarrow$ positive introspection on K
$KKp \wedge K\neg Kp$
$\Rightarrow$
$K(Kp \wedge \neg Kp)$
$\Rightarrow$
$K\bot$
$\Rightarrow$ given seriality (beliefs are consistent)
$\bot$

The Moore-sentence $p \wedge \neg Kp$

$K(p \wedge \neg Kp)$ is inconsistent—another proof.

$K(p \wedge \neg Kp)$
$\Rightarrow$
$Kp \wedge K\neg Kp$
$\Rightarrow$ property of belief derivable from introspection
$Kp \wedge \neg Kp$
$\Rightarrow$
$\bot$

The Moore sentence as an unsuccessful update

The Moore sentence as an unsuccessful update

You don't know that there was ice in Sevilla in 2012!

The Moore sentence as an unsuccessful update

You don't know that there was ice in Sevilla in 2012!

So now you do...

The Moore sentence as an unsuccessful update

You don't know that there was ice in Sevilla in 2012!

So now you do...

You don't know that there was ice in Sevilla in 2012!

The Moore sentence as an unsuccessful update

You don't know that there was ice in Sevilla in 2012!

So now you do...

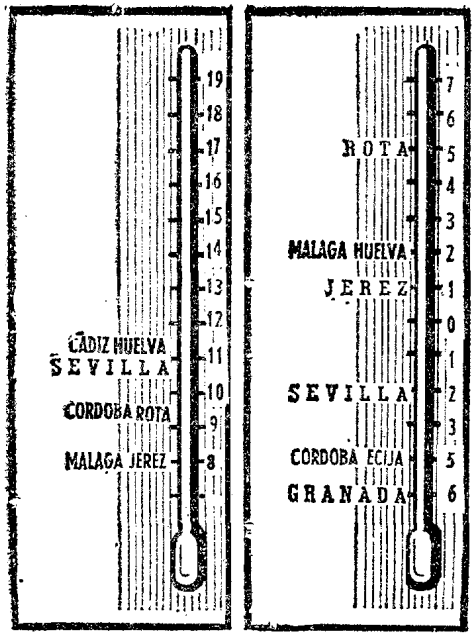
You don't know that there was ice in Sevilla in 2012! is a lie.

EL TIEMPO

2,4 BAJO CERO EN SEVILLA

Seguirá el viento Norte

MAXIMA MINIMA



El lunes igualamos la mínima del siglo para diciembre. Ayer la superamos al alcanzar nada menos que dos grados y medio bajo cero en Sevilla. La marca estaba, como indicábamos, en dos grados negativos.

Es posible que en algunas informaciones no figure esa cifra para la mínima, ya que ésta se produjo más tarde de lo habitual y no se reflejaba, por tanto, en la primera observación de la mañana, pero los dos grados y medio se alcanzaron, y constan ya en la climatología de nuestro aeropuerto.

una revuelta atmosférica que, aunque no traiga mucha agua, haga cambiar de signo la situación. Urge una suavización de las temperaturas en todas partes, pero, sobre todo, en Levante, porque los cuatro grados negativos registrados en Valencia pueden ser ya trágicos para aquella región, y no digamos lo que ocurriría de prolongarse aún más las actuales circunstancias.—R. CARBAJAL.

Información facilitada por el observatorio de San Pablo, a las dieciséis horas del día de ayer, para la región andaluza:

Información general: Los cielos se mantuvieron despejados o escasamente nubosos en toda la región, y las temperaturas fueron muy bajas, alcanzándose la cifra récord en el aeropuerto de Sevilla, en la época, de 2,4 bajo cero. Sopló terral moderado y en ocasiones fuerte en la costa mediterránea.

Tiempo probable (predicción válida hasta las dieciocho horas del día 13): Nubosidad variable, más bien escasa, en toda la región, con temperaturas bajas pero con tendencia a aumentar. Los vientos continuarán de componente Norte.

Le media de la presión barométrica ha sido de 764,8 mm.

Cinco grados bajo cero en Ecija

Ecija 12. Hoy se ha sentido en Ecija durante todo el día un frío intensísimo, especialmente en las primeras horas de la mañana, registrando el termómetro cinco grados bajo cero, según nos informa el Servicio Meteorológico local. Sin duda alguna, ha sido la temperatura más baja conocida hasta ahora.—Corresponsal.

PORTUARIAS

Sevilla

The Moore sentence as an unsuccessful update

You don't know that there was ice in Sevilla in 2012!

- By conversational implicature the announcement means: 'there was ice in Sevilla in 2012 and you don't know that.'
- Let proposition p stand for 'there was ice in Sevilla in 2012'.
- Write K for the epistemic modal operator modelling your knowledge. (Not mine!)
- 'there was ice in Sevilla in 2012' is formalized as $p \wedge \neg Kp$.
- There is no absurdity whatsoever.

The Moore sentence as an unsuccessful update

You don't know that there was ice in Sevilla in 2012!

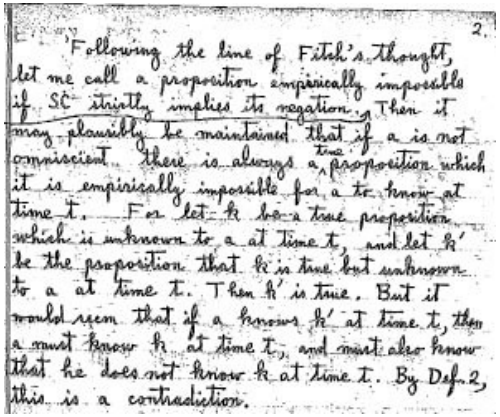
There is no absurdity whatsoever.

- $p \wedge \neg Kp$ is true before the announcement
- Kp is true after my announcement, and therefore also:
- $\neg p \vee Kp$
  $\neg p \vee \neg\neg Kp$
  $\neg(p \wedge \neg Kp)$
  the negation of the announcement

Fitch paradox

Following the line of Fitch's thought, let me call a proposition empirically impossible if SC strictly implies its negation. Then it may plausibly be maintained that if a is not omniscient there is always a true proposition which it is empirically impossible for a to know at time t. For let h be a true proposition which is unknown to a at time t, and let h' be the proposition that h is true but unknown to a at time t. Then h' is true. But it would seem that if a knows h' at time t, then a must know h at time t, and must also know that he does not know h at time t. By Def. 2, this is a contradiction.

Fitch paradox



2.

Following the line of Fitch's thought, let me call a proposition empirically impossible if SC strictly implies its negation. Then it may plausibly be maintained that if a is not omniscient there is always a time t and a proposition which it is empirically impossible for a to know at time t. For let k be a true proposition which is unknown to a at time t, and let k' be the proposition that k is true but unknown to a at time t. Then k' is true. But it would seem that if a knows k' at time t, then a must know k at time t, and must also know that he does not know k at time t. By Def. 2, this is a contradiction.

Alonzo Church. *First Anonymous Referee Report on Fitch's 'A Definition of Value'.* January or February 1945. (To Ernest Nagel, co-editor of the Journal of Symbolic Logic.)

Frederic B. Fitch. *A logical analysis of some value concepts.* The Journal of Symbolic Logic, 28(2):135–142, 1963.

Fitch paradox – knowability

Fitch's paradox is that some unknown truths are unknowable: $\exists p(p \wedge \neg Kp)$ is inconsistent with $\forall q(q \rightarrow \Diamond Kq)$.

$q \rightarrow \Diamond Kq$ for all q
$\Rightarrow$ for $q = p \wedge \neg Kp$
$(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$
$\Rightarrow$ on condition that $p \wedge \neg Kp$
$\Diamond K(p \wedge \neg Kp)$
$\Rightarrow$ assuming some reasonable semantics for $\Diamond$...
$K(p \wedge \neg Kp)$
$\Rightarrow$ as before
$\bot$

Successful and knowable

- 'postulate of success': after revision with φ, φ is believed.

  Successful formulas: $\varphi \to \langle !\varphi \rangle K\varphi$ is valid $\quad$ ($[!\varphi]K\varphi$ is valid)
  If φ is true, then after announcing φ, φ is known.

- 'Fitch's knowability': if φ is true, φ is knowable.

  Knowable formulas: $\varphi \to \Diamond K\varphi$ is valid
  If φ is true, then there is an announcement after which φ is known.

Fitch's 'paradox': not all formulas are knowable.
(namely not $p \wedge \neg Kp$)

Not all formulas are successful.
(namely not $p \wedge \neg Kp$)

Successful and knowable — example

Propositional variable p is knowable and is successful.

$p \to \langle !p \rangle Kp$
$p \to \Diamond Kp$

$$\underline{1} \text{——} 0 \quad \overset{!p}{\Rightarrow} \quad \underline{1}$$

$\langle !p \rangle Kp$ $\qquad$ p, Kp
$\Diamond Kp$

Successful and knowable — example

But formula $p \wedge \neg Kp$ is not knowable and is not successful.

$(p \wedge \neg Kp) \to \langle !(p \wedge \neg Kp) \rangle K(p \wedge \neg Kp)$ is invalid
$(p \wedge \neg Kp) \to \Diamond K(p \wedge \neg Kp)$ is invalid

$$\underline{1} \text{——} 0 \quad \overset{!(p \wedge \neg Kp)}{\Rightarrow} \quad \underline{1}$$

$p \wedge \neg Kp$ $\qquad$ $p, Kp, \neg p \vee Kp, \neg(p \wedge \neg Kp)$

Arbitrary announcement logic

Language $\varphi ::= p \mid \neg\varphi \mid (\varphi_1 \wedge \varphi_2) \mid K_a\varphi \mid [!\varphi_1]\varphi_2 \mid \Box\varphi$

Structures pointed Kripke models with epistemic accessibility relations $\rightarrow_a$ for each agent

Semantics

$$
\begin{array}{lll}
M, s \models p & \text{iff} & s \in V_p \\
M, s \models \neg\varphi & \text{iff} & M, s \not\models \varphi \\
M, s \models \varphi \wedge \psi & \text{iff} & M, s \models \varphi \text{ and } M, s \models \psi \\
M, s \models K_a\varphi & \text{iff} & \text{for all } t \in S : s \rightarrow_a t \text{ implies } M, t \models \varphi \\
M, s \models [!\varphi]\psi & \text{iff} & M, s \models \varphi \text{ implies } M|\varphi, s \models \psi \\
M, s \models \Box\varphi & \text{iff} & \text{for all epistemic } \psi : M, s \models [!\psi]\varphi
\end{array}
$$

$M|\varphi$: restriction of model M to the states where formula φ is true.
Abbreviations: $\Diamond\varphi$ for $\neg\Box\neg\varphi$, $\hat{K}_a\varphi$ for $\neg K_a\neg\varphi$, $\langle!\varphi\rangle\psi$ for $\neg[!\varphi]\neg\psi$.

Example of the semantics: $\Diamond(K_a p \vee K_a \neg p)$ is valid

$\Diamond\varphi$ is true in a model, iff
there is an epistemic ψ such that $\langle !\psi\rangle\varphi$ is true, iff
there is a ... model restriction such that φ is true in the restriction.

$\underline{1}$——0 $\Rightarrow$ $\underline{1}$

$!p$

$\Diamond(K_a p \vee K_a \neg p), \langle !p\rangle(K_a p \vee K_a \neg p)$ $p, K_a p$

1——$\underline{0}$ $\Rightarrow$ $\underline{0}$

$!\neg p$

$\Diamond(K_a p \vee K_a \neg p), \langle !\neg p\rangle(K_a p \vee K_a \neg p)$ $\neg p, K_a \neg p$

Moore-sentence:
$p \wedge \neg K_a p$

$!(p \wedge \neg K_a p)$

$\Rightarrow$ $K_a p, \neg(p \wedge \neg K_a p)$

Validities, theory

- $\models \Box(\varphi \wedge \psi) \leftrightarrow (\Box\varphi \wedge \Box\psi)$
- $\models \Box\varphi \rightarrow \varphi$
- $\models \Box\varphi \rightarrow \Box\Box\varphi$
- $\models \Diamond\Box\varphi \rightarrow \Box\Diamond\varphi$

– more expressive than epistemic logic
– complete finitary axiomatization
– non-compact
– undecidable [French & vDitm, AiML 2008]
– model checking PSPACE-complete [Ågotnes et al., JAL 2009]

[Balbiani, Baltag, v Ditmarsch, Herzig, Hoshi, de Lima 2007 & 08]
What can we achieve by arbitrary announcements? TARK 2007
'Knowable' as 'known after an announcement.' RSL 2008

Successful formulas and knowable formulas

- Positive: $\varphi ::= p|\neg p|\varphi_1 \vee \varphi_2|\varphi_1 \wedge \varphi_2|K_a\varphi|[!\neg\varphi_1]\varphi_2|\Box\varphi$
- Preserved: $\models \varphi \to \Box\varphi$
- Successful: $\models [!\varphi]\varphi$
- Knowable: $\models \varphi \to \Diamond K_a\varphi$

- Positive formulas are preserved. (And v.v.) Inductive case: $M, s \models [!\neg\varphi]\psi$ iff ($M, s \models \varphi$ or $M|\neg\varphi, s \models \psi$)
- Preserved formulas are successful. $\models \varphi \to \Box\varphi$ implies $\models \varphi \to [!\varphi]\varphi$ iff $\models [!\varphi]\varphi$.
- Successful formulas are knowable. $\models [!\varphi]\varphi$ iff $\models \varphi \to \langle!\varphi\rangle K_a\varphi$ implies $\models \varphi \to \Diamond K_a\varphi$.
- Some successful formulas are not positive: $\neg K_a p$.
- Some knowable formulas are not successful: $K_a(p \wedge \neg K_b p)$.

Successful formulas

- Syntactic characterization of single-agent successful formulas: [Holliday & Icard, AiML 2010].
- Commonly known formulas are successful: $C_A\varphi$.
- Syntactic characterization of multi-agent successful formulas?

Knowable formulas

Four ways of defining knowable as a logical validity:

1. $(\varphi \to \Diamond K\varphi) \wedge (\neg\varphi \to \Diamond K\neg\varphi)$
2. $\varphi \to \Diamond K\varphi$
3. $(\varphi \to \Diamond K\varphi) \vee (\neg\varphi \to \Diamond K\neg\varphi)$
4. $\Diamond K\varphi \vee \Diamond K\neg\varphi$

$(\varphi \to \Diamond K\varphi) \vee (\neg\varphi \to \Diamond K\neg\varphi)$
$\Leftrightarrow$
$\neg\varphi \vee \Diamond K\varphi \vee \varphi \vee \Diamond K\neg\varphi$
$\Leftrightarrow$
$(\varphi \vee \neg\varphi) \vee (\Diamond K\varphi \vee \Diamond K\neg\varphi)$
$\Leftrightarrow$
$\top$

- th-Knowable: $\models \varphi \to \Diamond K\varphi$
- wh-Knowable: $\models \Diamond K\varphi \vee \Diamond K\neg\varphi$

Everything is knowable

- $p \wedge \neg Kp$ is wh-knowable:
  if truthfully announced, you know that it is false.
- Every formula is wh-knowable!

  Proof sketch:
  Given M, s, announce the value of every prop. variable in φ.
  In the model restriction $M|\mathrm{var}(\varphi)$, φ is valid or $\neg\varphi$ is valid.
  In $M|\mathrm{var}(\varphi)$, $K\varphi$ is valid or $K\neg\varphi$ is valid.
  Therefore, $M, s \models \Diamond K\varphi \vee \Diamond K\neg\varphi$.

What is the axiomatization of knowability logic?
(the language without public announcements)

Knowability overview

Temporal aspects For every true proposition:

- ... we can get to know that it is true — *False*
- ... we can get to know that it was true — *False*
- ... we can get to know whether it is true — *True*
- ... we can get to know whether it was true — *False*

Multiagent aspects

- knowledge transfer: $K_a\varphi \to \Diamond K_b\varphi$
- knowledge diffusion: $D_A\varphi \to \Diamond C_A\varphi$

Non-public actions

- knowledge transfer: $\varphi \to \Diamond K_a\varphi$
  where $\Diamond$ quantifies over non-public actions.

Planning

- Whether $\models \varphi \to \Diamond\psi$ for initial conditions φ and goal ψ.

Some additional references

- van Ditm, Kooi, *The Secret of My Success*, Synthese 2006
- Gerbrandy, *The surprise examination in dynamic epistemic logic*, Synthese 2007
- van Benthem, *What one may come to know*, Analysis 2004
- van Ditmarsch, Ruan, Verbrugge, *Sum and Product in Dynamic Epistemic Logic*, JLC 2007
- Holliday, Icard, *Moorean Phenomena in Epistemic Logic*, AiML 2010
- van Ditmarsch, van der Hoek & Iliev, *Everything is knowable*, Theoria 2012
- Cordón, van Ditmarsch, Nepomuceno, *Dynamic Consequence and Announcement*, Review of Symbolic Logic, 2013
- Bozzelli, van Ditmarsch, French, Hales, Pinchinat, *Refinement Modal Logic*, Information & Computation, 2014