PHIL 222
Philosophical Foundations of Computer Science
Week 11, Thursday

Nov. 7, 2024

Epistemology (2)
Multi-Agent Systems and AI:
The "Problem of Logical Omniscience"
(cont'd)

Fagin–Halpern 1988:

"α-is-aware-of-φ",

Fagin–Halpern 1988:

"α-is-aware-of-φ", which can be interpreted variously:

- "α is aware of φ",
- "α is able to figure out the truth of φ",
- "α is able to compute the truth of φ within time T".

Fagin–Halpern 1988:

"α-is-aware-of-φ", which can be interpreted variously:

- "α is aware of φ",
- "α is able to figure out the truth of φ",
- "α is able to compute the truth of φ within time T".

We can then entertain various inference rules: e.g.,

- "α-is-aware-of-φ" $\iff$ "α-is-aware-of-(not-φ)",
- "α-is-aware-of-(φ-and-ψ)" $\implies$ "α-is-aware-of-φ",
- "α-is-aware-of-(φ-and-ψ)" $\iff$ "α-is-aware-of-(ψ-and-φ)".

Fagin–Halpern 1988:

"α-is-aware-of-φ", which can be interpreted variously:

- "α is aware of φ",
- "α is able to figure out the truth of φ",
- "α is able to compute the truth of φ within time T".

We can then entertain various inference rules: e.g.,

- "α-is-aware-of-φ" $\iff$ "α-is-aware-of-(not-φ)",
- "α-is-aware-of-(φ-and-ψ)" $\implies$ "α-is-aware-of-φ",
- "α-is-aware-of-(φ-and-ψ)" $\iff$ "α-is-aware-of-(ψ-and-φ)".

Then define "α explicitly knows that φ"

$$\iff (\alpha\text{-knows-that-}\varphi)\text{-and-}(\alpha\text{-is-aware-of-}\varphi).$$

Fagin–Halpern 1988:

"α-is-aware-of-φ", which can be interpreted variously:

- "α is aware of φ",
- "α is able to figure out the truth of φ",
- "α is able to compute the truth of φ within time T".

We can then entertain various inference rules: e.g.,

- "α-is-aware-of-φ" $\iff$ "α-is-aware-of-(not-φ)",
- "α-is-aware-of-(φ-and-ψ)" $\implies$ "α-is-aware-of-φ",
- "α-is-aware-of-(φ-and-ψ)" $\iff$ "α-is-aware-of-(ψ-and-φ)".

Then define "α explicitly knows that φ"

$$\iff (\alpha\text{-knows-that-}\varphi)\text{-and-}(\alpha\text{-is-aware-of-}\varphi).$$

So, given any logical truth φ, α knows that φ, but maybe only implicitly and not explicitly, since α may not be aware of φ.

To provide "α-is-aware-of-φ" with a truth condition,
we add "sets of accessible formulas" to the partition model.

To provide "α-is-aware-of-φ" with a truth condition,
we add "sets of accessible formulas" to the partition model.

In the partition model, we had:

- There are several states in each of which atomic sentences may be true or false.

To provide "α-is-aware-of-φ" with a truth condition,
we add "sets of accessible formulas" to the partition model.

In the partition model, we had:

- There are several states in each of which atomic sentences may be true or false.

Now we expand the above as follows:

- Several states, and each state w comes with two pieces of data:
  a. the set of atomic sentences that are true in w;

To provide "α-is-aware-of-φ" with a truth condition,
we add "sets of accessible formulas" to the partition model.

In the partition model, we had:

- There are several states in each of which atomic sentences may be true or false.

Now we expand the above as follows:

- Several states, and each state w comes with two pieces of data:
  a. the set of atomic sentences that are true in w;
  b. for each agent α, the set A^{α}_{w} of formulas she is aware of in w.

To provide "α-is-aware-of-φ" with a truth condition,
we add "sets of accessible formulas" to the partition model.

In the partition model, we had:

- There are several states in each of which atomic sentences may be true or false.

Now we expand the above as follows:

- Several states, and each state w comes with two pieces of data:
  a. the set of atomic sentences that are true in w;
  b. for each agent α, the set A^{α}_{w} of formulas she is aware of in w.

Then

"α-is-aware-of-φ" is true in w $\iff$ the formula "φ" is in the set A^{α}_{w}.

We can then entertain various restrictions on A^{α}_{w} — e.g.,

- "φ" is in $A^{\alpha}_{w} \iff$ "not-φ" is in A^{α}_{w},
- "φ-and-ψ" is in $A^{\alpha}_{w} \implies$ "φ" is in A^{α}_{w},
- "φ-and-ψ" is in $A^{\alpha}_{w} \iff$ "ψ-and-φ" is in A^{α}_{w}.

We can then entertain various restrictions on A^{α}_{w} — e.g.,

- "φ" is in A^{α}_{w} $\iff$ "not-φ" is in A^{α}_{w},
- "φ-and-ψ" is in A^{α}_{w} $\implies$ "φ" is in A^{α}_{w},
- "φ-and-ψ" is in A^{α}_{w} $\iff$ "ψ-and-φ" is in A^{α}_{w}.

Then

"α-explicitly-knows-that-φ" is true in w

$\iff$ "φ" is true in all the α-cell mates of w and is in A^{α}_{w}.

But the use of sets of accessible formulas may be susceptible of Stalnaker's criticism (of what he calls a "storage model"):

But the use of sets of accessible formulas may be susceptible of Stalnaker's criticism (of what he calls a "storage model"):

> There are at least two different notions of implicit belief, a broad notion and a narrow notion. On the broad notion, the implicit beliefs of a believer include everything the believer is committed to in virtue of having the explicit beliefs he has [. . .]. This will include all the deductive consequences of the explicit beliefs [. . .]. [I]mplicit beliefs are by definition deductively closed, for ordinary believers as well as for those with extraordinary computational powers. The claim that something is implicitly believed says nothing about whether the believer has access to that belief — whether the believer will assert or assent to it, or act as if he thinks it is true. No one thinks that implicit belief in this broad sense is an analysis of belief in the ordinary sense [. . .]. [I]mplicit belief in this sense tells us no more than explicit belief about the inferential powers of the believer. [Stalnaker (1991), "The Problem of . . . , I", p. 434]

We saw these with Sili, whose knowledge base contains

i. The Burj Khalifa is a building with a height of 830 m.
ii. The Burj Khalifa is the tallest building in the world.

We saw these with Sili, whose knowledge base contains

i. The Burj Khalifa is a building with a height of 830 m.

ii. The Burj Khalifa is the tallest building in the world.

a. You: How tall is the Burj Khalifa?

We saw these with Sili, whose knowledge base contains

i. The Burj Khalifa is a building with a height of 830 m.

ii. The Burj Khalifa is the tallest building in the world.

a. You: How tall is the Burj Khalifa?

Not-silly-Sili: 830 m.

We saw these with Sili, whose knowledge base contains

i. The Burj Khalifa is a building with a height of 830 m.

ii. The Burj Khalifa is the tallest building in the world.

a. You: How tall is the Burj Khalifa?

Not-silly-Sili: 830 m.

b. You: What is the height of the Burj Khalifa times the biggest known prime number's next prime number?

We saw these with Sili, whose knowledge base contains

i. The Burj Khalifa is a building with a height of 830 m.

ii. The Burj Khalifa is the tallest building in the world.

a. You: How tall is the Burj Khalifa?

Not-silly-Sili: 830 m.

b. You: What is the height of the Burj Khalifa times the biggest known prime number's next prime number?

Not-silly-Sili: Am I supposed to know that???

We saw these with Sili, whose knowledge base contains

i. The Burj Khalifa is a building with a height of 830 m.

ii. The Burj Khalifa is the tallest building in the world.

a. You: How tall is the Burj Khalifa?

Not-silly-Sili: 830 m.

b. You: What is the height of the Burj Khalifa times the biggest known prime number's next prime number?

Not-silly-Sili: Am I supposed to know that???

c. You: How tall is the tallest building in the world?

We saw these with Sili, whose knowledge base contains

(i) The Burj Khalifa is a building with a height of 830 m.

(ii) The Burj Khalifa is the tallest building in the world.

(a) You: How tall is the Burj Khalifa?

Not-silly-Sili: 830 m.

(b) You: What is the height of the Burj Khalifa times the biggest known prime number's next prime number?

Not-silly-Sili: Am I supposed to know that???

(c) You: How tall is the tallest building in the world?

Not-silly-Sili: 830 m (by (i) and (ii)).

The problem is that the distinction between implicit and explicit belief is being used to do two different jobs that one distinction is not suited to do. The manifest fact that we are not logically omniscient is a fact about our computational limitations — the fact that some of the information that is implicit in what we know or believe is, because of computational limitations, not accessible to us. To get at belief and knowledge in the ordinary sense we need a distinction between what is accessible and what is implicit but inaccessible. The explicit-implicit distinction is sometimes tacitly assumed to be this distinction. But, if it is, then it is a completely different distinction from the one that the storage model makes between two different forms in which information is represented, the distinction between propositions expressed by sentences written down in the belief box and propositions not written down there but somehow implicit in the ones that are. [pp. 435f.]

We may put Stalnaker's criticism this way:

We may put Stalnaker's criticism this way:

What we need is an account of how, from (a), an agent can obtain (c) but not all of (b).

(a) explicit knowledge / belief;

(b) implicit knowledge / belief by deductive closure;

(c) implicit but accessible knowledge / belief.

We may put Stalnaker's criticism this way:

What we need is an account of how, from (a), an agent can obtain (c) but not all of (b).

(a) explicit knowledge / belief;

(b) implicit knowledge / belief by deductive closure;

(c) implicit but accessible knowledge / belief.

The storage model (or the awareness approach) has (b), "knows-that". But which of (a) and (c) is the storage (the set A_w^α) supposed to express?

We may put Stalnaker's criticism this way:

What we need is an account of how, from (a), an agent can obtain (c) but not all of (b).

(a) explicit knowledge / belief;

(b) implicit knowledge / belief by deductive closure;

(c) implicit but accessible knowledge / belief.

The storage model (or the awareness approach) has (b), "knows-that". But which of (a) and (c) is the storage (the set A_w^α) supposed to express?

- If it is (a), then how does the model distinguish (c) from (b)?

We may put Stalnaker's criticism this way:

What we need is an account of how, from (a), an agent can obtain (c) but not all of (b).

(a) explicit knowledge / belief;

(b) implicit knowledge / belief by deductive closure;

(c) implicit but accessible knowledge / belief.

The storage model (or the awareness approach) has (b), "knows-that". But which of (a) and (c) is the storage (the set A_w^α) supposed to express?

- If it is (a), then how does the model distinguish (c) from (b)?
- If it is (c), then how does the model distinguish (c) from (a)?

We may put Stalnaker's criticism this way:

What we need is an account of how, from (a), an agent can obtain (c) but not all of (b).

(a) explicit knowledge / belief;

(b) implicit knowledge / belief by deductive closure;

(c) implicit but accessible knowledge / belief.

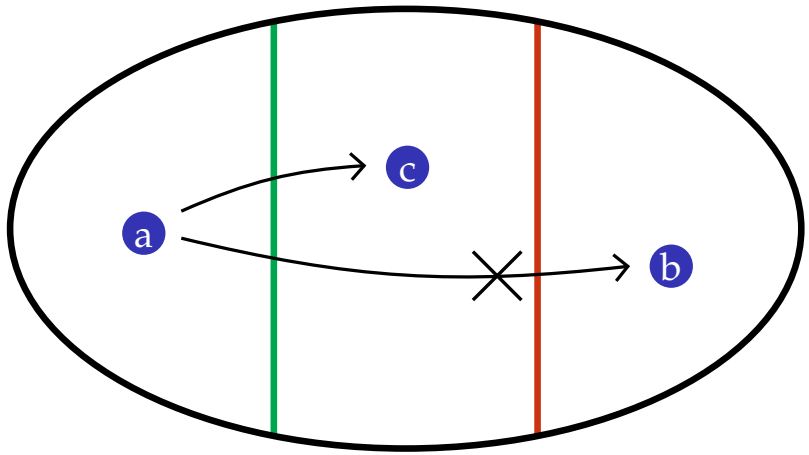
The storage model (or the awareness approach) has (b), "knows-that". But which of (a) and (c) is the storage (the set A_w^α) supposed to express?

- If it is (a), then how does the model distinguish (c) from (b)?
- If it is (c), then how does the model distinguish (c) from (a)?

Either way, it is left unexplained how an agent can obtain (c) but not all of (b) from (a).

In abstract terms (not recommendable for your writing exercise!),

- we need two lines separating the three levels of knowledge / belief, because we need to explain why an agent can cross the first line but not the second,



- but the storage model only draws one line (whichever of the two needed lines it may be).

Ordinary knowledge is a capacity, and ordinary belief a disposition. Because of our computational limitations, we may have the capacity constituted by the knowledge that P, or the disposition constituted by the belief that P, while at the same time lacking the capacity or disposition that we would have if we knew or believed some deductive consequence of P. But what is the capacity or disposition as capacity or disposition to do? The storage model has nothing to say about this, and so has little promise of clarifying the problem of logical omniscience. [p. 436]

2. Questions help turn implicit knowledge explicit.

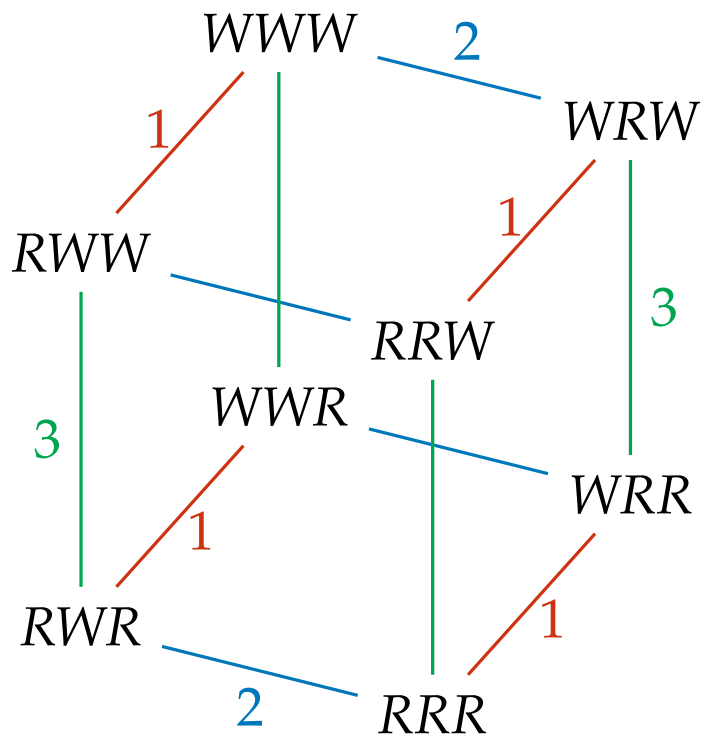
Some researchers pursue this idea.

2. Questions help turn implicit knowledge explicit.

Some researchers pursue this idea.

For instance, in the partition-model approach . . .

- Remember the puzzle of the hats. Announcements are modelled by dynamic updates of the model (deleting states, in particular).

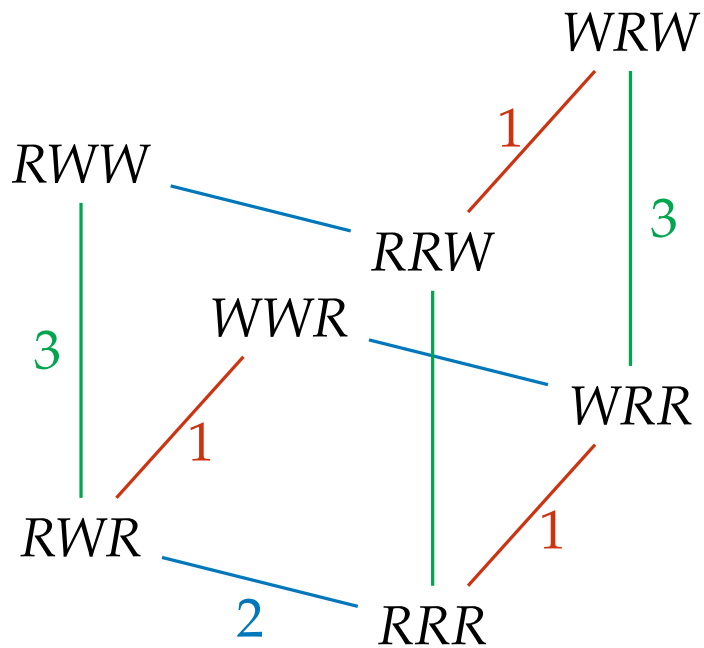


2. Questions help turn implicit knowledge explicit.

Some researchers pursue this idea.

For instance, in the partition-model approach . . .

- Remember the puzzle of the hats. Announcements are modelled by dynamic updates of the model (deleting states, in particular).

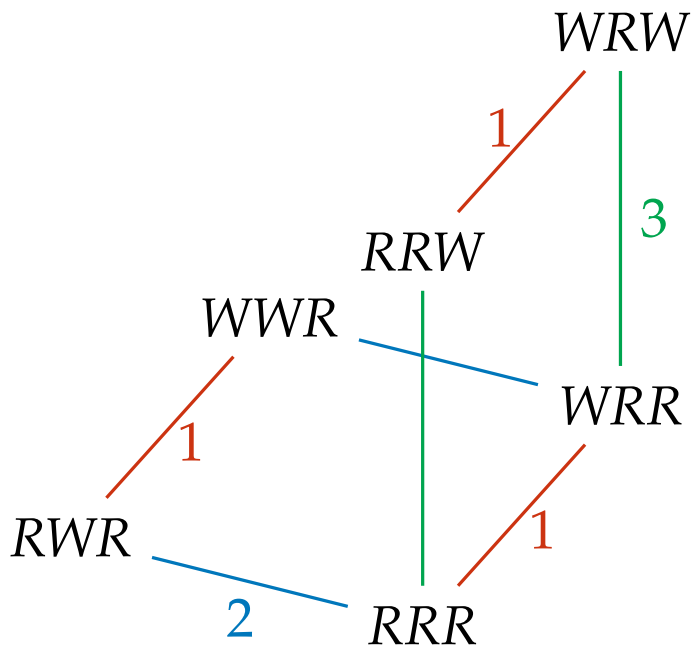


2. Questions help turn implicit knowledge explicit.

Some researchers pursue this idea.

For instance, in the partition-model approach . . .

- Remember the puzzle of the hats. Announcements are modelled by dynamic updates of the model (deleting states, in particular).

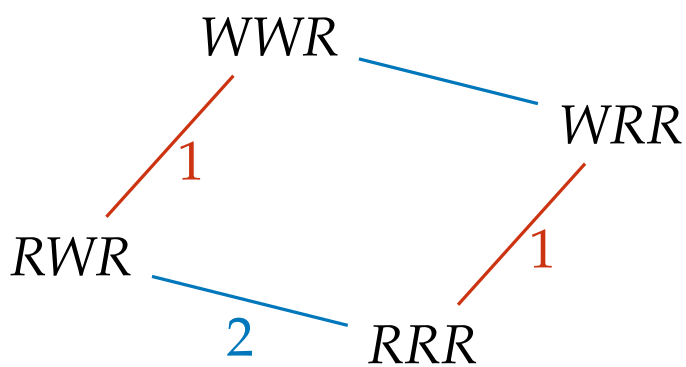


2. Questions help turn implicit knowledge explicit.

Some researchers pursue this idea.

For instance, in the partition-model approach . . .

- Remember the puzzle of the hats. Announcements are modelled by dynamic updates of the model (deleting states, in particular).



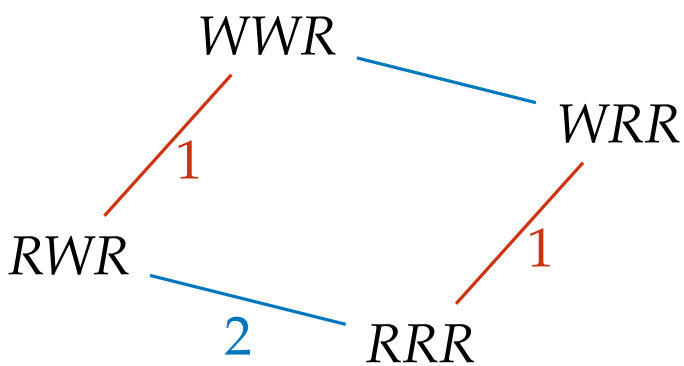
2. Questions help turn implicit knowledge explicit.

Some researchers pursue this idea.

For instance, in the partition-model approach . . .

- Remember the puzzle of the hats. Announcements are modelled by dynamic updates of the model (deleting states, in particular).

Similarly, questions can be modelled by updates of the model.



2. Questions help turn implicit knowledge explicit.

Some researchers pursue this idea.

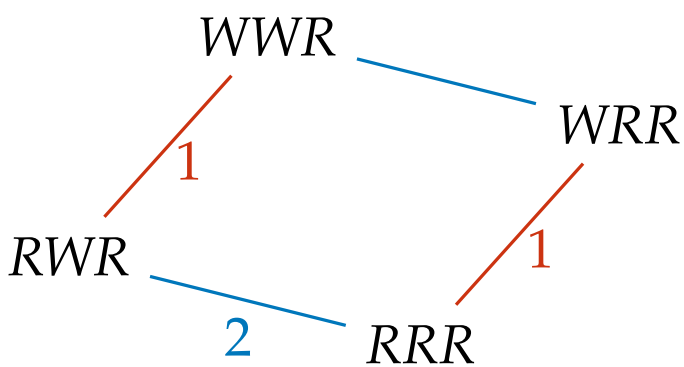
For instance, in the partition-model approach . . .

- Remember the puzzle of the hats. Announcements are modelled by dynamic updates of the model (deleting states, in particular).

Similarly, questions can be modelled by updates of the model.

There are several proposals as to exactly what updates should model questions — but e.g.,

- The question "Is φ true or not?" to an agent α makes her aware of φ (or not), i.e. it places φ (and $\neg\varphi$) in A^{α}_{w} in the new model.



- The question "Is φ true or not?" to an agent α makes her aware of φ (or not), i.e. it places φ (and $\neg\varphi$) in A^{α}_{w} in the new model.

- The question "Is φ true or not?" to an agent α makes her aware of φ (or not), i.e. it places φ (and $\neg\varphi$) in A^{α}_{w} in the new model.

  This then updates the situation (i) into (ii).

  (i) "α-knows-that-φ" is true but "α-is-aware-of-φ" is not — i.e., α knows that φ implicitly but not explicitly.

  (ii) "α-knows-that-φ" and "α-is-aware-of-φ" are both true — i.e., α knows that φ not only implicitly but also explicitly.

- The question "Is φ true or not?" to an agent α makes her aware of φ (or not), i.e. it places φ (and $\neg\varphi$) in A^{α}_{w} in the new model.

  This then updates the situation (i) into (ii).

  (i) "α-knows-that-φ" is true but "α-is-aware-of-φ" is not — i.e., α knows that φ implicitly but not explicitly.

  (ii) "α-knows-that-φ" and "α-is-aware-of-φ" are both true — i.e., α knows that φ not only implicitly but also explicitly.

One may then e.g. incorporate questions into the protocol description in the puzzle of the hats.

- The question "Is φ true or not?" to an agent α makes her aware of φ (or not), i.e. it places φ (and $\neg\varphi$) in A^{α}_{w} in the new model.

  This then updates the situation (i) into (ii).

  (i) "α-knows-that-φ" is true but "α-is-aware-of-φ" is not — i.e., α knows that φ implicitly but not explicitly.

  (ii) "α-knows-that-φ" and "α-is-aware-of-φ" are both true — i.e., α knows that φ not only implicitly but also explicitly.

One may then e.g. incorporate questions into the protocol description in the puzzle of the hats.

This and other proposals may have issues, but let's assume we have a good model of questions. We may then narrow down "actions" to quenstion-answering and try the following criterion of accessibility:

(A) α knows that φ in the accessible way

$\iff$ α can easily answer if asked a question whose content is φ.

- Ⓐ α knows that φ in the accessible way

  $\Longleftrightarrow$ α can easily answer if asked a question whose content is φ.

- (A) α knows that φ in the accessible way
  $\iff$ α can easily answer if asked a question whose content is φ.

But Stalnaker finds several problems.

For instance, take "$43 \times 37 = 1591$" for φ.

A α knows that φ in the accessible way

$\iff$ α can easily answer if asked a question whose content is φ.

But Stalnaker finds several problems.

For instance, take "$43 \times 37 = 1591$" for φ. Suppose we are asked

a "Is it true that 43 times 37 equals 1591?"

A. α knows that φ in the accessible way

$\iff$ α can easily answer if asked a question whose content is φ.

But Stalnaker finds several problems.

For instance, take "$43 \times 37 = 1591$" for φ. Suppose we are asked

a. "Is it true that 43 times 37 equals 1591?"

b. "What are the prime factors of 1591?"

(A) α knows that φ in the accessible way

$\iff$ α can easily answer if asked a question whose content is φ.

But Stalnaker finds several problems.

For instance, take "$43 \times 37 = 1591$" for φ. Suppose we are asked

(a) "Is it true that 43 times 37 equals 1591?"

(b) "What are the prime factors of 1591?"

We can easily answer (a), but not (b), although the two questions have exactly the same content, φ (i.e. $43 \times 37 = 1591$).

(A) α knows that φ in the accessible way

$\iff$ α can easily answer if asked a question whose content is φ.

But Stalnaker finds several problems.

For instance, take "$43 \times 37 = 1591$" for φ. Suppose we are asked

(a) "Is it true that 43 times 37 equals 1591?"

(b) "What are the prime factors of 1591?"

We can easily answer (a), but not (b), although the two questions have exactly the same content, φ (i.e. $43 \times 37 = 1591$). So,

- (a) brings out our knowledge that φ, whereas (b) does not.

Ⓐ α knows that φ in the accessible way

$\Longleftrightarrow$ α can easily answer if asked a question whose content is φ.

But Stalnaker finds several problems.

For instance, take "$43 \times 37 = 1591$" for φ. Suppose we are asked

ⓐ "Is it true that 43 times 37 equals 1591?"

ⓑ "What are the prime factors of 1591?"

We can easily answer ⓐ, but not ⓑ, although the two questions have exactly the same content, φ (i.e. $43 \times 37 = 1591$). So,

- ⓐ brings out our knowledge that φ, whereas ⓑ does not.
- But, according to Ⓐ, do we have this knowledge accessible or not? It seems that Ⓐ + ⓐ says Yes, whereas Ⓐ + ⓑ says No.

Ⓐ α knows that φ in the accessible way

$\Longleftrightarrow$ α can easily answer if asked a question whose content is φ.

But Stalnaker finds several problems.

For instance, take "$43 \times 37 = 1591$" for φ. Suppose we are asked

ⓐ "Is it true that 43 times 37 equals 1591?"

ⓑ "What are the prime factors of 1591?"

We can easily answer ⓐ, but not ⓑ, although the two questions have exactly the same content, φ (i.e. $43 \times 37 = 1591$). So,

- ⓐ brings out our knowledge that φ, whereas ⓑ does not.
- But, according to Ⓐ, do we have this knowledge accessible or not? It seems that Ⓐ + ⓐ says Yes, whereas Ⓐ + ⓑ says No.

This seems to mean that "α knows that φ" has no truth value independently of questions.

Ⓐ α knows that φ in the accessible way

$\Longleftrightarrow$ α can easily answer if asked a question whose content is φ.

But Stalnaker finds several problems.

For instance, take "$43 \times 37 = 1591$" for φ. Suppose we are asked

ⓐ "Is it true that 43 times 37 equals 1591?"

ⓑ "What are the prime factors of 1591?"

We can easily answer ⓐ, but not ⓑ, although the two questions have exactly the same content, φ (i.e. $43 \times 37 = 1591$). So,

- ⓐ brings out our knowledge that φ, whereas ⓑ does not.
- But, according to Ⓐ, do we have this knowledge accessible or not? It seems that Ⓐ + ⓐ says Yes, whereas Ⓐ + ⓑ says No.

This seems to mean that "α knows that φ" has no truth value independently of questions. One may take this to show, e.g.,

- We need algorithms and their complexity in the analysis.

Ⓐ α knows that φ in the accessible way

$\Longleftrightarrow$ α can easily answer if asked a question whose content is φ.

But Stalnaker finds several problems.

For instance, take "$43 \times 37 = 1591$" for φ. Suppose we are asked

ⓐ "Is it true that 43 times 37 equals 1591?"

ⓑ "What are the prime factors of 1591?"

We can easily answer ⓐ, but not ⓑ, although the two questions have exactly the same content, φ (i.e. $43 \times 37 = 1591$). So,

- ⓐ brings out our knowledge that φ, whereas ⓑ does not.
- But, according to Ⓐ, do we have this knowledge accessible or not? It seems that Ⓐ + ⓐ says Yes, whereas Ⓐ + ⓑ says No.

This seems to mean that "α knows that φ" has no truth value independently of questions. One may take this to show, e.g.,

- We need algorithms and their complexity in the analysis.
- "Knows *that*" is derivative from other notions of knowlege.

Ethics (1):
Introduction

Apple's 'sexist' credit card investigated by US regulator

11 November 2019

A US financial regulator has opened an investigation into claims Apple's credit card offered different credit limits for men and women.

It follows complaints - including from Apple's co-founder Steve Wozniak - that algorithms used to set limits might be inherently biased against women.

New York's Department of Financial Services (DFS) has contacted Goldman Sachs, which runs the Apple Card.

Any discrimination, intentional or not, "violates New York law", the DFS said.

The Bloomberg news agency reported on Saturday that tech entrepreneur David Heinemeier Hansson had complained that the Apple Card gave him 20 times the credit limit

https://www.nytimes.com/2019/11/11/technology/artificial-intelligence-bias.html

We Teach A.I. Systems Everything, Including Our Biases

Researchers say computer systems are learning from lots and lots of digitized books and news articles that could bake old attitudes into new technology.

By Cade Metz

Nov. 11, 2019

SAN FRANCISCO — Last fall, Google unveiled a breakthrough artificial intelligence technology called BERT that changed the way scientists build systems that learn how people write and talk.

But BERT, which is now being deployed in services like Google's internet search engine, has a problem: It could be picking up on biases in the way a child mimics the bad behavior of his parents.

BERT is one of a number of A.I. systems that learn from lots and lots of digitized information, as varied as old books, Wikipedia entries and news articles. Decades and even centuries of biases — along with a few new ones — are probably baked into all that material.

What is the field of ethics about?

Ethics has a couple branches (read Darwall!):

What is the field of ethics about?

Ethics has a couple branches (read Darwall!):

1. **Normative ethics.** "Substantive" normative questions like "What is valuable?", "What is morally obligatory?", etc.

2. **Meta-ethics.** "Concerned with more abstract philosophical issues that underlie" substantive normative questions. Intersects with philosophy of language (e.g. "concerning the meaning and content of ethical judgments"), philosophy of mind, metaphysics, and epistemology. Also, moral psychology.

What is the field of ethics about?

Ethics has a couple branches (read Darwall!):

1. **Normative ethics.** "Substantive" normative questions like "What is valuable?", "What is morally obligatory?", etc.
   a. **Applied ethics.** Specific ethical issues and cases.
   b. **Normative theory.** "[P]rinciples, concepts, and ideals that can be cited in support of ethical judgments about cases".
2. **Meta-ethics.** "Concerned with more abstract philosophical issues that underlie" substantive normative questions. Intersects with philosophy of language (e.g. "concerning the meaning and content of ethical judgments"), philosophy of mind, metaphysics, and epistemology. Also, moral psychology.

What is the field of ethics about?

Ethics has a couple branches (read Darwall!):

1. **Normative ethics.** "Substantive" normative questions like "What is valuable?", "What is morally obligatory?", etc.
   a. **Applied ethics.** Specific ethical issues and cases.
   b. **Normative theory.** "[P]rinciples, concepts, and ideals that can be cited in support of ethical judgments about cases".
2. **Meta-ethics.** "Concerned with more abstract philosophical issues that underlie" substantive normative questions. Intersects with philosophy of language (e.g. "concerning the meaning and content of ethical judgments"), philosophy of mind, metaphysics, and epistemology. Also, moral psychology.

a and b interact a *lot* (discussed shortly).

People discuss exactly how (or whether) 1 and 2 interact.

Computer ethics.

The subfield called "**computer ethics**" spans normative ethics and meta-ethics.

1. What is good, bad, right, wrong, just or unjust in computing and the use of computers and ICTs (information-and-communication technologies), or more generally in society in "the information age". E.g., what (special) responsibilities computing professionals have.
2. Foundational questions pertaining to entities found in computing. E.g., can an AI be a moral agent (that has responsibilities and/or rights)?

— although there may be some questions in the intersection: e.g., when an AI has caused harm, who is responsible, the AI or human beings involved in its creation and/or usage?

Computer ethics.

The subfield called "**computer ethics**" spans normative ethics and meta-ethics.

1. What is good, bad, right, wrong, just or unjust in computing and the use of computers and ICTs (information-and-communication technologies), or more generally in society in "the information age". E.g., what (special) responsibilities computing professionals have.
2. Foundational questions pertaining to entities found in computing. E.g., can an AI be a moral agent (that has responsibilities and/or rights)?

— although there may be some questions in the intersection: e.g., when an AI has caused harm, who is responsible, the AI or human beings involved in its creation and/or usage?

The chapters of this course on ethics concern questions of type 1.

One topic we discuss is software property rights — e.g. whether or how we should protect them.

One topic we discuss is software property rights — e.g. whether or how we should protect them.

We read Johnson (with Miller), *Computer Ethics*, 4th ed., Chapter 5.

"Scenario 5.2":

> Bingo Software Systems has an idea for a new file organizing system that Bingo believes will be significantly more intuitive than existing systems. Bingo is a small company employing twenty people. It obtains venture capital and spends three years developing the system. Over the course of the three years, Bingo invests approximately two million dollars in development of the system. When completed, the new system is successfully marketed for about a year, and Bingo recovers about 50 percent of its investment. However, after the first year, several things happen that substantially cut into sales. First, [. . .]

First, a competing company, Pete's Software, starts distributing a file organizing system that performs many of the same functions that Bingo's software does, but has a different interface and a few new features. Pete's Software has put its software on its website for free download using a GPLv2 license (this is called "Free Software"). Pete's Software hopes to recoup its investment by selling its services customizing the software for individual clients. It appears that Pete's programmers studied Bingo's system, adopted a similar general approach, and then produced a new piece of software that provided functionality comparable to that in Bingo's software but more efficiently. As far as Bingo programmers can tell, Pete's programmers did not copy any of the source or object code of the Bingo system. Instead, it seems that Pete's software was newly built, that is, from the ground up.

According to its lawyer, Bingo would be unlikely to prevail in a copyright or a "look and feel" lawsuit against Pete's Software, and extended legal proceedings would be prohibitively expensive for a small company like Bingo. Customers, primarily small businesses, appear to be downloading Pete's software and then making multiple copies for internal use. Some of those companies hire Pete's Software to help them and many don't. But Pete's Software has plenty of business, whereas Bingo's business seems to be slipping away. Bingo is unable to recover the full costs of developing its original system, and within a few years files for bankruptcy. Pete's Software, pleased by its success, begins another project in which they target another market segment currently served by proprietary software; they plan to again develop a Free Software alternative.

Is this situation unfair? Has Pete's Software wronged Bingo Software Systems?

Philosophers' approach.

This is a question of computer ethics + applied ethics. How do we tackle such questions?

Philosophers' approach.

This is a question of computer ethics + applied ethics. How do we tackle such questions?

Instead of the term "applied ethics" (which may suggest that applied ethics is to normative theory what applied math is to pure math), Darwall proposes "case ethics", like case law.

> Just as there is "case law," the findings of judges about the issues brought before them, including, crucially, the reasoning or *ratio* that led to their conclusions, so also is there case ethics: our considered judgments about specific ethical issues or cases along with the reasons or principled reflections that underlie our judgments.
>
> [pp. 17f.]

On the one hand, judgments on cases invoke normative theories:

> [W]e commit ourselves implicitly to some theory (or range of theories) whenever we give reasons to support our judgments.
>
> [p. 18]

On the other, the study of normative theory often invokes cases, too:

> [T]heories are often formulated and evaluated by reflecting on the ethically relevant features of cases. Thus some philosophers maintain that we can appreciate the general moral relevance of a distinction between killing and letting die (or, more generally yet, between causing evils and letting them happen) by reflecting on a specific case like Judith Thomson's famous "trolley problem," [. . .].
>
> [p. 17]

Specific ethical judgments (case ethics) depend essentially on normative theory — due to their essential reason-dependence.

> [W]e commit ourselves implicitly to some theory (or range of theories) whenever we give reasons to support our judgments. What's more, there is a sense in which we commit ourselves to the existence of some justifying background theory whenever we even *make* an ethical judgment. This is because of an important feature of ethical concepts and properties that we might call their reason- or warrant-dependence.

Specific ethical judgments (case ethics) depend essentially on normative theory — due to their essential reason-dependence.

> [W]e commit ourselves implicitly to some theory (or range of theories) whenever we give reasons to support our judgments. What's more, there is a sense in which we commit ourselves to the existence of some justifying background theory whenever we even *make* an ethical judgment. This is because of an important feature of ethical concepts and properties that we might call their reason- or warrant-dependence. When, for example, I judge that something is good, I say, not just that I value it, but that there is reason to value it — that valuing it is warranted, an attitude one ought to have.

Specific ethical judgments (case ethics) depend essentially on normative theory — due to their essential reason-dependence.

> [W]e commit ourselves implicitly to some theory (or range of theories) whenever we give reasons to support our judgments. What's more, there is a sense in which we commit ourselves to the existence of some justifying background theory whenever we even *make* an ethical judgment. This is because of an important feature of ethical concepts and properties that we might call their reason- or warrant-dependence. When, for example, I judge that something is good, I say, not just that I value it, but that there is reason to value it — that valuing it is warranted, an attitude one ought to have. As a logical matter, however, this can be true only if something has *other* properties: the reasons for valuing it. And such reasons cannot simply consist in the property that it is good, since that is itself the property of there being such reasons. [pp. 18f.]

Unlike, say, the property of yellowness, which might attach to something all by itself, as it were, ethical properties require, by their very nature, completion by further properties that are their reasons or grounds. If I judge a certain experience to be valuable, I must think it has aspects that make it good, features that are the grounds of its value. Or if I think that a certain action is morally required, I must think there are certain characteristics of the action and the situation that make it morally obligatory, features that are the grounds of its obligatoriness. And these thoughts commit me to the existence of background normative theories. I am committed to thinking there are truths that relate an experience's having certain properties to its value, such that any experience that had exactly those (and no other ethically relevant) properties would be valuable also, other things being equal. Or, similarly, I am committed to thinking there exists some valid moral principle that relates an action's having certain features to its being morally required. [p. 19]

This is why, e.g., when discussing software property rights — whether or how we should protect them — one typically needs to

i. identify what characterizes properties that should be protected by property rights, as well as figuring out principles behind property rights (why there are property rights, why we ought to protect them, etc.),
ii. judge whether software has these characteristics.

This is why, e.g., when discussing software property rights — whether or how we should protect them — one typically needs to

i. identify what characterizes properties that should be protected by property rights, as well as figuring out principles behind property rights (why there are property rights, why we ought to protect them, etc.),
ii. judge whether software has these characteristics.

In this chapter of the course, we discuss software property rights and laws currently protect them, but for the reason above we will also review two relevant normative theories:

a. utilitarianism,
b. a natural rights argument, by John Locke in particular.