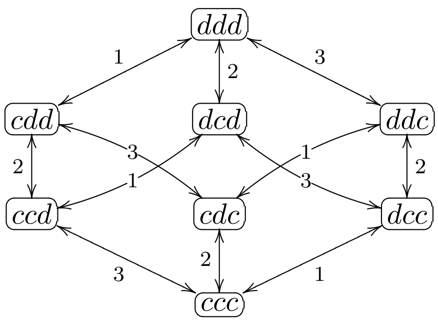
Answers to Warming-Up Exercises

1. (a) Represent all the facts, as well as all the children's knowledge, in a Kripke model $\mathbf{M}$ with three agents $\mathcal{A} = \{1, 2, 3\}$ and three atomic propositions $Prop = \{d_1, d_2, d_3\}$ (where d_i means "child i is dirty"). Either draw the model, or else specify formally: the set W of possible worlds; the knowledge relations R_1, R_2, R_3; the valuation. Also, specify which world is the real world, and denote that world by w.

   ANSWER: Use atomic sentences d_i to denote that "*child i is dirty*". 3 children: so *8 possible worlds*. We may use c_i ("i is clean") as an abbreviation for $\neg d_i$. The valuation of each world is made obvious by the way we encode the world, e.g. $\nu(ddc) = \{d_1, d_2\}$ The **real** world is $w = (ddc)$, satisfying $d_1 \wedge d_2 \wedge \neg d_3$.

   We are representing by arrows the children's **knowledge** relations, so we'll get an **epistemic** model: all relations $R_1, R_2, R3$ are **equivalence relations**. So in particular they are **reflexive**, but for simplicity of drawing **I skipped the loops**:



   We will denote this model by $\mathbf{M}$.

   (b) Is the above model $\mathbf{M}$ an *epistemic* model?

   ANSWER: Yes (as already explained above), all relations are equivalence relations, so $\mathbf{M}$ is an epistemic model.

   (c) Check (using the semantics of K_i) that (in the real world w) *everybody knows that at least one of them is dirty*; i.e. show (using the semantics of K_i) that

$$w \models_{\mathbf{M}} K_1(d_1 \vee d_2 \vee d_3) \wedge K_2(d_1 \vee d_2 \vee d_3) \wedge K_3(d_1 \vee d_2 \vee d_3)$$

ANSWER: To check that

$$w \models_{\mathbf{M}} K_1(d_1 \vee d_2 \vee d_3) \wedge K_2(d_1 \vee d_2 \vee d_3) \wedge K_3(d_1 \vee d_2 \vee d_3)$$

we use the semantics of K: we need to check that all the worlds reachable from w in one step by either R_1-arrows or R_2-arrows or R_3-arrows satisfy $d_1 \vee d_2 \vee d_3$. But the only world **not** satisfying $d_1 \vee d_2 \vee d_3$ is (ccc), and this world is **not** reachable in one step from $w = (ddc)$. Hence, **all** the worlds reachable in one step from w **do** indeed satisfy $d_1 \vee d_2 \vee d_3$.

(d) Is it common knowledge (in the real world w) that at least one of them is dirty? Prove your answer formally by encoding it as a sentence in epistemic logic and checking that it is true at the world w in model $\mathbf{M}$ (using the semantics of the common knowledge operator).

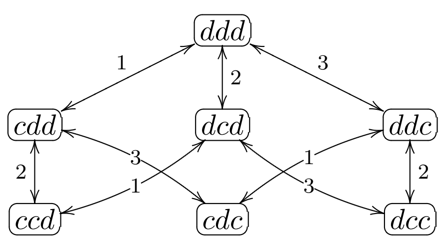
ANSWER: No, it is **not** common knowledge that at least of one of them is dirty. To check that

$$w \not\models_{\mathbf{M}} Ck(d_1 \vee d_2 \vee d_3)$$

we use the semantics of the common knowledge operator Ck (which is just another notation for $C\Box$ in epistemic models). For $Ck(d_1 \vee d_2 \vee d_3)$ to be true at w, we would need that **all** worlds reachable from w by any concatenation of arrows (of any length) satisfy $d_1 \vee d_2 \vee d_3$. But this is **not** the case: the world (ccc) is reachable from w by concatenations of arrows (since $w = (ddc)R_1(cdc)R_2(ccc)$) though it does not satisfy $d_1 \vee d_2 \vee d_3$ (since $(ccc) \models \neg d_1 \wedge \neg d_2 \wedge \neg d_3$). Hence, $Ck(d_1 \vee d_2 \vee d_3)$ is **not** true at w.

(e) Father makes a *truthful public announcement*: "At least one of you is dirty". Draw (or specify formally) the *updated model* $\mathbf{M}'$ representing the children's knowledge **after** this announcement.

ANSWER: After the public announcement $!(d_1 \vee d_2 \vee d_3)$, the updated model $\mathbf{M}'$ is:



(f) The scenario continues as in last week's slides: in response to Father's first question ("Do you know if you are dirty or not, and if so which?"), children answer truthfully, publicly and simultaneously. What will they answer (in the real world w)? Justify your answer, by writing down one long formal sentence φ_1 (in the language of epistemic logic) that expresses all the children's simultaneous answers to this first question (in the real world w), and checking that this sentence is true at w in the model $\mathbf{M}'$.

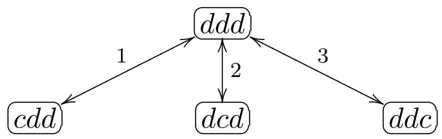
ANSWER: They all answer "I don't know if I am dirty or not". The formal sentence encoding all the answers (i.e. "nobody knows if he is dirty or not") is:

$$\neg K_1 d_1 \wedge \neg K_1 \neg d_1 \wedge \neg K_2 d_2 \wedge \neg K_2 \neg d_2 \wedge \neg K_3 d_3 \wedge \neg K_3 \neg d_3.$$

It is easy to check that this sentence is indeed *true at the actual world* $w = (ddc)$: each of the children has some uncertainty arrow going from this world $w = (ddc)$ to some other world, in which his/her own state is different than in w (i.e. clean instead of dirty, or dirty instead of clean).

(g) Let us interpret children's simultaneous answering to Father's first question as a truthful public announcement $!\varphi_1$ of the sentence you wrote in the previous part. What is the updated model $\mathbf{M}''$ after this new public announcement? (Draw, or else formally specify, this model.)

ANSWER: After the public announcement of the sentence written in the previous part ($\neg K_1 d_1 \wedge \neg K_1 \neg d_1 \wedge \neg K_2 d_2 \wedge \neg K_2 \neg d_2 \wedge \neg K_3 d_3 \wedge \neg K_3 \neg d_3$), the new updated model $\mathbf{M}''$ is obtained by *deleting* all the worlds (from the model in part (e)) in which this sentence was *false*, i.e. in which one or another of the children *knows* (s)he dirty or knows (s)he clean. Those are exactly the worlds (ccd), (cdc) and (dcc). So, by deleting those worlds we obtain the new model $\mathbf{M}''$:



(h) If Father repeats the same question *now* (after the children's public answering of the first question), what will the children answer now? Justify your answer, by writing down one long formal sentence φ_2 (in the language of epistemic logic) that expresses all the children's simultaneous answers to this second question (at the real world w), and checking that this sentence is true at w in the model $\mathbf{M}''$.

ANSWER: The dirty children will answer "I know I am dirty", while the clean child will answer "I don't know". We can encode these answers in the following sentence φ_2:

$$K_1 d_1 \wedge K_2 d_2 \wedge \neg K_3 d_3 \wedge \neg K_3 \neg d_3$$

To justify that this is the right answer, we need to check that

$$w \models_{\mathbf{M}''} K_1 d_1 \wedge K_2 d_2 \wedge \neg K_3 d_3 \wedge \neg K_3 \neg d_3$$

To show this, we use the semantics of K: in the model the only world accessible from $w = (ddc)$ by R_1-arrows is w itself, which does satisfy d_1, so that (all worlds reachable from w by R_1-arrows satisfy d_1, hence) w satisfies $K_1 d_1$. The argument for child 2 is similar. As for child 3, both $w = (ddc)$ and (ddd) are accessible from w by R_3-arrows, so since (ddc) satisfies $\neg d_3$ while (ddd) satisfies d_3, child 3 doesn't know (i.e. we have $w \models \neg K_3 d_3 \wedge \neg K_3 \neg d_3$).

(i) Show that in the next round of questioning, the clean child will answer "I know I am clean". Justify your answer, by drawing the new model $\mathbf{M}'''$ after the previous step (i.e. after all children answered the second question) and checking that

$$w \models_{\mathbf{M}'''} K_3 \neg d_3.$$

ANSWER: To compute the result of the new update $!\varphi_2$, we have to once again *delete* from the previous model $\mathbf{M}''$ all the worlds in which the announced sentence φ_2 was *false*. Those worlds are: (cdd) (since child 1 has an uncertainty arrow between this world and (ddd), so he doesn't know if he's dirty or not, i.e. $K_1 d_1$ is false in this world, contrary to the announcement φ_2), (dcd) (since child 2 has an uncertainty arrow between this world and (ddd), hence $K_2 d_2$ is false in this world, again contrary to the announcement φ_2) and (ddd) (since child 1 has an uncertainty arrow between (ddd) and (cdd), and child 2 has an uncertainty arrow between (ddd) and (dcd), hence neither child 1 nor child 2 know that they are dirty, contrary to the announced statement φ_2).

By deleting from$\mathbf{M}''$ these three worlds, we obtain a model $\mathbf{M}'''$ consisting of **only one world** $w = (ddc)$**, with loops for all agents**.

Obviously, since $\neg d_3$ is true at all worlds in this model (since there is only one world (ddc)), we have that

$$w \models_{\mathbf{M}'''} K_3 \neg d_3$$

2. Prove the validity of the "Veracity of Knowledge" $K_a \varphi \Rightarrow \varphi$ on epistemic models.

PROOF: let $\mathbf{M} = (W, R_a, \ldots, \nu)$ be any epistemic model, and let $w \in W$ be any world in it.

To prove our claim, suppose that $K_a\varphi$ is true at w, i.e.

$$(1) \qquad w \models K_a\varphi,$$

and we need to **prove** that: (?) $w \models \varphi$.

For this, note that, by the semantics of K_a $(= \Box_a)$, (1) implies that

$$(1') \qquad \forall v.\ wR_av \Rightarrow (v \models \varphi).$$

But $\mathbf{M}$ is an epistemic model so (by the reflexivity of R_a) we have wR_aw. So we can take in particular $v := w$ in (1'), and conclude that

$$w \models \varphi.$$

DONE!

3. Prove the validity of "Positive Introspection of Knowledge" $K_a\varphi \Rightarrow K_aK_a\varphi$ on epistemic models.

   PROOF:

   Let $\mathbf{M} = (W, R_a, \ldots, \nu)$ be an epistemic model and $w \in W$ be any world. To prove our claim, suppose that

   $$(1) \qquad w \models K_a\varphi.$$

   Let v be *any arbitrary world such* wR_av and s be *any arbitrary world such* vR_as. By transitivity, we have wR_as, so applying (1) and the semantics of K_A, we get

   $$s \models \varphi.$$

   Since this is for ANY world s with vR_as, we conclude that

   $$v \models K_a\varphi.$$

   Once again, this conclusion holds for ANY world v with wR_av, hence:

   $$w \models K_aK_a\varphi. \qquad (\text{ DONE! })$$

4. Prove the validity of the left-to-right implication in the "Fixed Point Axiom" (slide 64) on epistemic models.

   PROOF: let $\mathbf{M} = (W, R_a, \ldots, \nu)$ be any model and $w \in W$ be any world such that

   $$(1) \qquad w \models_{\mathbf{M}} Ck\varphi.$$

Unfolding this assumption (using the semantics of Ck) we get

(2) $w' \models_{\mathbf{M}} \varphi$ for every finite chain $w = w_o R_{a_1} w_1 \ldots R_{a_k} w_k = w'$.

From this, we need to **prove** that:

$$(?) \quad w \models_{\mathbf{M}} (\varphi \wedge K_1 Ck\varphi \wedge \ldots \wedge K_n Ck\varphi).$$

To prove this, it is enough to show two things:

$$(??) \quad w \models_{\mathbf{M}} \varphi, \qquad \text{and}$$

$$(???) \quad w \models_{\mathbf{M}} K_i Ck\varphi \qquad \text{for all } i \in \{1, \ldots, n\}.$$

To show (??), just note that $wR_a w$, so by taking in particular $w' := w$ in (2) we obtain (??).

To show (???), let v be *any arbitrary world such* $wR_i v$, and let v' *any world reachable from* v *by any finite chain* $v = v_0 R_{a_1} v_1 \ldots R_{a_k} v_k = v'$. Then v' is also reachable *from* w by the finite chain $wR_i v R_{a_1} v_1 \ldots R_{a_k} v_k = v'$. So, by (2), we must have $v' \models_{\mathbf{M}} \varphi$.

So this conclusion holds for *ANY world* v' *reachable from* v *by any finite chain.* Hence (by the semantics of Ck), we have

$$v \models_{\mathbf{M}} Ck\varphi.$$

But v was just *ANY arbitrary world such* $wR_i v$. So (by the semantics of K_i), we obtain (???).

5. Prove the validity of the "Induction Axiom" (slide 64) on epistemic models.

   PROOF: To prove the **left-to-right implication of the Induction Axiom**:

   let $\mathbf{M} = (W, R_a, \ldots, \nu)$ be any model and $w \in W$ be any world such that

   $$(1) \qquad w \models_{\mathbf{M}} Ck(\varphi \Rightarrow K_1\varphi \wedge \ldots \wedge K_n\varphi).$$

   We need to **prove** that

   $$(?) \quad w \models_{\mathbf{M}} (\varphi \Rightarrow Ck\varphi).$$

   To show this, we assume that we have

   $$(2) \qquad w \models_{\mathbf{M}} \varphi,$$

   and we need to **prove** that

   $$(??) \quad w \models_{\mathbf{M}} Ck\varphi.$$

But (by the semantics of Ck) this is the same as proving that:

(???) $\quad w' \models_{\mathbf{M}} \varphi$ for every finite chain $w = w_o R_{a_1} w_1 \ldots R_{a_k} w_k = w'$.

We now proceed to prove (???) **by induction on the length k of the finite chain**:

Step $k = 0$: For chains of length $k = 0$ this just says that

$$w' \models_{\mathbf{M}} \varphi \text{ for } w = w_0 = w',$$

which is the same as $w \models_{\mathbf{M}} w$: so this just follows directly from (2).

Inductive Step: Assume (???) to be *true for all chains of some given length k* starting from w. We need to *prove it for chains of length* $k+1$.

For this, let $w = w_o R_{a_1} w_1 \ldots R_{a_k} w_k R_{a_{k+1}} w'$ be such a chain of length $k+1$ starting from w. Then, by our inductive hypothesis, it follows that

(3) $\quad w_k \models_{\mathbf{M}} \varphi \quad$ (since w_k is reachable from w by a chain of length k).

From (1) (using the semantics of Ck), we also have that all worlds reachable by finite chains from w satisfy $(\varphi \Rightarrow K_1\varphi \wedge \ldots \wedge K_n\varphi)$. Hence in particular (since w_k is reachable from w by such a chain) we have:

$$(4) \quad w_k \models_{\mathbf{M}} (\varphi \Rightarrow K_1\varphi \wedge \ldots \wedge K_n\varphi).$$

Putting (3) and (4) together, we conclude by Modus Ponens that

$$w_k \models_{\mathbf{M}} (K_1\varphi \wedge \ldots \wedge K_n\varphi)$$

and hence in particular

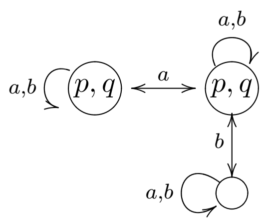
$$w_k \models_{\mathbf{M}} K_{a_{k+1}}\varphi.$$

From this, using the semantics of K and the fact that $w_k R_{a_{k+1}} w'$, we conclude that

$$w' \models_{\mathbf{M}} \varphi.$$

But w' is here ANY arbitrary world reachable from w by any chain of length $k+1$. So we proved our inductive step.

6. Consider the following model with three possible worlds and two agents a and b:



In the (p, q)-world on the left, both a and b know both p and q (i.e. $K_a p \wedge K_a q \wedge K_b p \wedge K_b q$ holds in this world), but $p \wedge q$ is NOT common knowledge. (In fact, in that same world, a doesn't know that b knows p, and a doesn't know that b knows q.)