COMPUTATIONAL LOGIC: FINAL 2017-18

Question 1. (*15 points*) Prove (via a *counterexample*) that

$$\neg\Box_a p \not\Rightarrow B_a \neg\Box_a p$$

Question 2. (*35 points*) The sensors of a self-driving car detect a human-shaped object, standing in the middle of the street. The car *believes* it's a human, and so it believes that it is NOT safe to drive ahead. (By "safe", the car means safe for human pedestrians. So it is "safe" for the car to drive ahead if and only if there are no humans in the middle of the street.) **In reality** (unknown to the car), the object is *just a mannequin* that somebody left standing in the middle of the street. IF the car was given the information that object is not a human, then the car would *believe (conditional on this information)* that the object *is just a mannequin* and that there is *no human* there (and so that it is in fact *safe* to drive ahead). **In reality**, in the back of the mannequin there is also a human (-a reckless pedestrian who is attempting to cross the street); but the car doesn't know this, since the human is completely covered by the mannequin. IF the car was given the information that there is a mannequin **and** a human behind it, then the car would *believe (conditional on this richer information)* that it is *NOT safe* to drive ahead. It is possible that the car's sensors are malfunctioning (-so that in fact there is *no mannequin and no human ahead*, and hence *it is safe* to drive), but the car considers this possibility to be the *most implausible* of all.

1. (*8 points*) *Represent* this situation as a *single-agent plausibility model* $\mathbf{M}_1$ (with the car as the agent), with 4 possible worlds. Use only 3 atomic sentences: m for "there is a mannequin on the street", h for "there is a human on the street, and s for "it is safe to drive ahead"

2. (*8 points*) The car deploys its infra-red sensors to detect if the observed object emits any heat, and concludes that this is not the case, i.e. the object must be a **mannequin**. The car strongly trusts its sensors, so it performs a **radical upgrade** with m. **Draw** a *plausibility model* for the situation **after** this upgrade.

3. (*8 points*) Somebody observing the scene from the side pavement shouts at the car "Beware, it is NOT safe to drive ahead!". The car has audio sensors, so it hears the announcement, and it weakly trusts the source, hence it performs a **conservative upgrade**. **Draw** a *plausibility model* for the situation **after** this upgrade. **What does the car believe now** (after the upgrade) about *whether or not there is a human ahead* (h)?

4. (*8 points*) Suppose instead that the events described in the last two parts happened in the reverse order: FIRST, the (weakly trusted) observer shouts "It is NOT safe to drive", and THEN the car deploys its (strongly trusted) infra-red sensors to conclude that the object is a mannequin (because it emits no heat). **Draw** a *plausibility model* for the situation *after this alternative scenario* (by performing on the initial model first a conservative upgrade then a radical upgrade). **What does the car believe now** (after this alternative scenario) about the *presence of humans* (h) and *safety of driving* (s)?

Question 3. (*53 points*) Each of two agents, Alexandru and Bob, has a natural number written on her forehead. Let $n_a \in \{0, 1, 2, \ldots\}$ be Alexandru's number and $n_b \in \{0, 1, 2, \ldots\}$ be Bob's number. It is **common knowledge** that: (i) *nobody can see his own number*, (ii) *each of them can see the other's number*; (iii) *one of the numbers is the immediate successor of the other (in any order)*: i.e. there are two cases, either $n_a = n_b + 1$ or $n_b = n_a + 1$; (iv) whenever any of them is in doubt about these two cases ($n_a = n_b + 1$, and $n_b = n_a + 1$), he considers them to be *equally plausible.*

1. (*29 points*) In all the subparts (a)-(e) of this part, we make the following background assumption: it is **common knowledge** that the two children **strongly trust** each other.

   (a). **Draw** a *plausibility model* $\mathbf{M}$, with two agents (a for Alexandru, b for Bob), to represent the above situation.

   (b). The following question is raised: "*Do you believe that your number is (strictly) larger than the other's number, do you believe that your number is smaller, or you believe neither of these* (i.e. consider the two possibilities to be *equally plausible*)? Alexandru can either answer $B_a(n_a > n_b)$ ("*I believe my number is larger*"), or $B_a(n_a < n_b)$ ("*I believe my number is smaller*"), or $\neg B_a(n_a > n_b) \wedge \neg B_a(n_a < n_b)$ ("*I believe none of the two*"). And similarly for Bob.

Alexandru is the FIRST to answer publicly: "*I believe none of the two*". **Draw** a *plausibility model* $\mathbf{M}'$ for the situation **after** this announcement (given our background assumption: *common knowledge of strong trust*).

(c). **Next**, it's Bob's turn to answer publicly the same question, and he says: "*I believe none of the two*". **Draw** a *plausibility model* $\mathbf{M}''$ for the situation *after* this.

(d). **Next**, it's again Alexandru's turn, and he says publicly: "*I now believe my number is larger than Bob's*". **Draw** a *plausibility model* $\mathbf{M}'''$ for the situation *after* this (given the same background assumption).

(e). Assuming that in fact **both agents were sincere** every time (i.e. *they told what they really believed*), **what is the real world**, i.e. **what are the numbers**?

2. (*24 points*) Let us consider an *alternative scenario*, that starts in the same initial conditions as above, but now *we change the background assumption*: it is now **common knowledge** that Alexandru **strongly distrusts** Bob *but* that Bob **strongly trusts** Alexandru.

   (a). Starting again from the initial situation, Alexandru answers the question first: "*I believe neither*". **Represent** this announcement as *an event plausibility model* $\mathbf{\Sigma}$ (given our *background assumption*: Bob strongly trusts Alexandru); and **draw** a *state plausibility model* $\mathbf{M}_1$ for the situation **after** this, by computing the Action-Priority Update of the original model $\mathbf{M}$ with $\mathbf{\Sigma}$.

   (c). **Next**, Bob also says: "*I believe none of the two*". **Represent** this announcement as *an event plausibility model* $\mathbf{\Sigma}_1$ (given our *background assumption*: Alexandru strongly **distrusts** Bob); and **draw** a *state plausibility model* $\mathbf{M}_2$ for the situation **after** this new round, by computing the Action-Priority Update of the state model $\mathbf{M}_1$ with $\mathbf{\Sigma}_1$.

   (d). *Finally*, it's Alexandru's turn again. Assuming that Alexandru is *sincere* (telling his real belief), and assuming the *real world is the same as the previous scenario* (i.e. the one in your answer to part (1e)), **what will he answer this time**? Is this (expressed) belief **true**?