COMPUTATIONAL LOGIC: Sample Final

IMPORTANT NOTE: This exam consists of 3 questions, *worth in total 100 points*. As the mid-term, the exam is **open-book**.

Question 1. (*15 points*)

1. (*8 points*) Prove *semantically* that: **if an agent believes ϕ then she strongly believes that this belief is safe**. In other words, show that

   $$B_a\phi \Rightarrow Sb_a\Box_a\phi$$

   is **valid** on plausibility models.

   HINT: You have to assume that all "best" worlds are ϕ-worlds, and prove two things: (a) there exist some $\Box_a\phi$-world(s); (b) every $\Box_a\phi$-world is strictly more plausible than every $\neg\Box_a\phi$-world. For (b), you may try to show it by assuming the opposite of the conclusion (i.e. that there exist some $\Box_a\phi$-world w and some $\neg\Box_a\phi$-world s with $w \leq_a s$) and deriving a contradiction.

2. (*7 points*) Prove (via a *counterexample*) that

   $$B_a\phi \not\Rightarrow \Box_a Sb_a\phi$$

   (where $\Box$ is safe belief, and Sb is strong belief).

   HINT: You need to give an example of a specific sentence ϕ (e.g. an atomic formula p), a specific model $\mathbf{M}$ and a world w in that model, such that w satisfies both $B_a\phi$ and $\neg\Box_a Sb_a\phi$. (You may use the validity $\neg Sb_a\phi \Rightarrow \neg\Box_a Sb_a\phi$, which follows from the validity $\neg P \Rightarrow \neg\Box_a P$, itself a consequence of the validity $\Box_a P \Rightarrow P$, mentioned in the lectures.)

Question 2. (*40 points*) After entering an Italian restaurant, Helen **doesn't know** for sure what's the quality of the service or of the food. Still, she **believes** both that the restaurant has good food and that it has good service. But, **if** it turns out that (*at least one of*) her beliefs are wrong, then **(conditional on this information) she'd still believe** that the food was good (but the service bad), rather than that the food was bad. However, **in case**

the food was bad, then (given this information) she'd rather believe that at least the service was good (rather than that both the food and service were bad).

In reality (unknown to Helen), the restaurant has bad food but good service.

1. (*6 points*) Represent Helen's beliefs, knowledge and conditional beliefs, using a **plausibility model M** with four possible worlds, and using the atomic sentences f for "the food is good" and s for "the service is good".

2. (*6 points*) Now Helen sees another customer angrily shouting at the waiter that he will never come back to this restaurant again! From this, Helen gathers that the other customer must be dissatisfied with either the service or the food (or both). Of course, Helen **doesn't know** for sure that this is the reason (maybe the customer simply thought that the bill was too high? or maybe the customer has a mental illness that makes him prone to unjustified fits of anger?). But she **strongly believes** that this is the reason, hence she performs a **radical upgrade** of her beliefs with the information that *either the food or the service (or both) are bad*.

   Represent Helen's new belief structure after this event as a **plausibility model M$'$**.

3. (*12 points*)(a) **What does Helen believe about the service now** (after the upgrade), and (b) **is this belief SAFE**? Justify your answers.

   HINT: You have to recall what the real world is!

4. (*8 points*) Soon after the incident described in the previous part, the waiter comes to Helen's table to take her order, he is very polite and in general behaves impeccably. From this, Helen gathers that the service might actually be good in this restaurant. She **believes** that the waiter's behavior is a proof in this sense, but her belief is rather **weak**; she **doesn't** strongly believe it: it's too soon to be sure, and besides she's still worried because of the incident she just witnessed. So Helen performs a **conservative upgrade** of her beliefs with the information that *the service is good*.

(a) **Represent** Helen's new belief structure (after this new upgrade) as a plausibility model $\mathbf{M}''$. (b) **What does Helen believe about the food after this**?

5. (*8 points*) Soon, the food comes and Helen takes a bite, and her worst expectations are now confirmed: **the food is bad**, no doubt about that! So she performs an **update** with this new information.

   (a) **Represent** Helen's new belief structure (after this last event) as a **plausibility model**.

   (b) *At the end* of the above scenario, **does Helen still have any false beliefs**? *Justify your answer* (giving either an example of false belief, or arguing why all her beliefs are true).

Question 3 (*45 points*): Alexandru and Beatrix are playing a game, with Christina as a referee. We consider Christina only as a **source of information**, NOT as an agent (-so you *don't have to worry about Christina's beliefs, or draw any plausibility arrows for her*). It is **common knowledge** that only 3 cards are left in the deck: Ace, King and Queen. Christina takes one card, looks at it privately, and places it face-down on the table (so that nobody else can see its face). So it is **common knowledge** that: *only Christina knows the card*, and that *the others only know it's one of the 3 cards listed above*. Moreover, it is also **common knowledge** that: both Alexandru and Beatrix *believe* (based on past experience) that *the card on the table is an Ace* (though they can't know this for sure); and that, *in case that their belief is wrong (i.e. conditional on the assumption that the card is NOT an Ace) they both believe that it'd be a King*.

1. (*5 points*) *Draw a multi-agent plausibility model* $\mathbf{M}_0$ for the above situatio, with two agents, a for Alexandru and b for Beatrix (-since Christina is NOT an agent here, but just a source of information), and three atomic sentences A, K and Q (where A means "*the card on the table is an Ace*" etc.).

2. (*10 points*) Now Christina (the referee) says publicly: "*The card on the table is NOT an Ace*". It is common knowledge that: Alexandru *strongly trusts* Christina (to tell the truth, so he has attitude $\Uparrow$ towards her), while Beatrix *strongly distrusts* her.

   Draw an **event plausibility model $\mathbf{\Sigma}$** for the above action.

3. (*10 points*) *Represent (draw) a model* $\mathbf{M}_1$ for the situation *after* the action described in the previous part, by computing the Action-Priority update $\mathbf{M}_1 = \mathbf{M}_0 \otimes \mathbf{\Sigma}$.

4. (*10 points*) We now change the previous scenario a bit. Christina makes exactly the same public announcement as above ("*The card on the table is not an Ace*"). But now it is common knowledge that: (1) each of the two players (Alexandru and Beatrix) *either strongly trusts* Christina ($\Uparrow$) or *strongly distrusts* her ($\Uparrow^-$); (2) each player *knows his/her attitude* towards Christina; (3) each of the two players *believes (without knowing for sure) that the other player strongly trusts* Christina.

   Represent this action as an **event plausibility model** $\mathbf{\Sigma}'$.

   HINT: This is a cube, with 8 possible actions.

5. (*10 points*) Starting again from the original situation (in part 1), suppose the action that we described in the previous part happens.

   Represent (draw) a model $\mathbf{M}'_1$ for the situation *after* this action, by computing the Action-Priority update $\mathbf{M}'_1 = \mathbf{M}_0 \otimes \mathbf{\Sigma}'$.