Backpropagation. Is backpropagation considered biologically plausible? If so, why, if not, why not?
1. No. In biological brains, backpropagation would lead to the vanishing gradient problem, in which gradients would go to zero before reaching the necessary depth in the network.
2. Yes. Like many AI principles, it is expressly inspired by known biological feedback connections.
3. **No. Backpropagation requires identical feedback connections for all feedforward connections.**
4. Yes. This is the role played by dopamine in the mammalian brain.

Grid Coding. We can use fMRI to find grid-like coding in the human brain, because:
1. When running along a hexagonal grid in a virtual reality environment, grid cells will respond periodically (at 4 Hz) whenever you cross the humps of a grid and this fast periodicity is what fMRI picks up.
2. We can't. As explained in class, grid cells are remapped from environment to environment, so this cannot be done in a reasonable fMRI experiment.
3. **Depending on your direction (modulo 60 degrees) in a space, you will either traverse the activations or silences of all grid cells. So, we analyze fMRI responses contingent on movement direction.**
4. We can use sophisticated machine learning techniques to find the signatures of grid-like coding, especially in entorhinal cortex.

Divisive Normalization population receptive field modeling. There are multiple parameters in the DN pRF model that represent constants in the numerator and denominator in its division:
1. Activation constant (B), in the numerator, determines the amount of response compression.
2. **Normalization constant (D), in the denominator, determines the amount of response compression.**
3. Normalization constant (D), in the numerator, determines the amount of surround inhibition.
4. **Activation constant (B), in the numerator, determines the amount of surround inhibition.**
5. Normalization constant (B), in the denominator, determines the amount of surround inhibition.

Divisive Normalization as a general principle. What 2 statements about divisive normalization are true?
1. **DN allows responses to be scaled to some background, for example it can rescale local neural responses to light based on the mean light level across the retina**
2. **DN can create winner-take-all suppression between neurons that represent different visual orientations, for example.**
3. DN can implement the temporal difference learning algorithm for reinforcement learning, but fails to capture the dynamics of the Rescorla-Wagner model
4. DN is one of the basic ingredients of recurrent neural network architectures, without which backpropagation-through-time cannot be performed.

Single-layer perceptron. There is a specific operation that a single-layer perceptron cannot perform. Which?
1. The AND operation
2. The NAND operation
3. The OR operation
4. **The XOR operation**
5. The single-layer perceptron is a universal function approximator: it can perform any operation

Grid Coding. How do responses in Entorhinal cortex (EC) lead to place cell responses in Hippocampus?
1. **Grid cells form a Fourier basis for the representation of space: place cells could be created through a linear combination of grid cell responses.**
2. EC already maintains a place cell code, and hippocampus inherits these place cell responses directly.
3. EC combines retinotopic activations throughout the brain to create a world-centric representation of space such as those in place cells.
4. EC maintains a history of where an animal has been in the past, and by integrating over all past locations EC can create knowledge of where an animal is now.

Replay. A population of place cells:
1. Can replay a sequence of events (a route) more slowly than it happened in actual life so that memory consolidation processes can lay down memories better.
2. Can lay down memories, but its activations during navigational decision-making are divorced from choices that an animal makes.

3. **Can replay a sequence of events (a route) faster than it happened in actual life so that memory consolidation processes connect distant cortical representations.**
4. Has place cells that are fixed to the environment, and they always have the same place field when we move from environment to environment.

Multi-layer perceptron. There is a specific operation that a multi-layer perceptron cannot perform. Which?
1. The AND operation
2. The NAND operation
3. The OR operation
4. The XOR operation
5. **The multi-layer perceptron is a universal function approximator: it can perform any operation**

Grid Coding. *Select the wrong statement.* Entorhinal cortex grid cells
1. In a single individual differ in their grid spacing (frequency)
2. In a single individual differ in their grid offset (phase)
3. **In a single individual differ in their grid orientation**
4. Are anchored to world-centered space

Tolman.
1. Was a behaviorist that, in 1924, discovered extinction in reinforcement learning
2. Was a behaviorist that, in 1932, discovered blocking in reinforcement learning, together with Roger Kamin
3. **Proved, in 1948, that rats maintain a cognitive map of their surroundings even without explicit reward**
4. Discovered grid cells together with May-Britt Moser, and received the nobel prize in 2014.

Encoding models. The space in which encoding models cannot be formulated:
1. Retinotopic visual space
2. Tonotopic auditory frequency
3. Semantic space
4. World-centric space
5. **There is no space in which encoding models cannot be formulated**

Biological Realism. *Find the wrong answer.* The basic perceptron model fails to mimic biological neurons in:
1. The spatial structure of the dendritic tree, that supports intricate integration of incoming information
2. The temporal delay structure of signals coming in on the dendritic tree, again supporting intricate integration of incoming information
3. The timing specificity of action potentials
4. **its fundamental input-output architecture**

Grid fit procedure. *Find the wrong answer.* Caveats of grid fitting:
1. It has limited precision (parameters are fixed on the grid)
2. It has predetermined range (meaning you can misspecify this and fail to find the global minimum)
3. For high-dimensional parameter spaces, the amount of possible grid elements explodes.
4. **Cannot be used for least-squares estimation, and has to be used for maximum likelihood estimation**

Gradient Descent. *Find the wrong answer.*
1. Gradient descent uses the same rule (delta rule) as that which drives reinforcement learning in for example the Rescorla-Wagner model.
2. Gradient descent determines the direction to step in based on the gradient $d/d\theta \, J(\theta)$ of our cost function at the present step
3. **Gradient Descent cannot be used for least-squares estimation, and has to be used for maximum likelihood estimation**
4. The step size in gradient descent is directly related to the learning rate in a reinforcement learning context.

Reinforcement learning. The learning rate in a Rescorla-Wagner model:
1. Determines the ceiling of the learning: the height of its asymptote
2. Depends on the size of the prediction error

3. Determines the speed at which learning tends towards its asymptote
4. Is larger for a larger reward.

Temporal Difference Learning TD solves this problem of the Rescorla-Wagner model:
1. RW cannot function in volatile environments due to its fixed learning rate
2. RW does not capture the explore-exploit trade-off
3. **RW only predicts immediate rewards: no higher-order conditioning**
4. RW doesn't incorporate actions, TD does

Signal detection theory. Imagine a simple signal-in-noise SDT experiment. In the first half, I reward hits but not correct rejections, and in the second half of the experiment, I reward correct rejections but not hits. What will the impact on my participants' behavior be?
1. They will be more sensitive in the first half of the experiment, and less sensitive in the second: this will impact D-prime.
2. **They will adjust their criterion, and be more conservative in the second half while they were more liberal in the first..**
3. Their percentage correct will be higher for 'noise' (no signal) trials in the second half of the experiment
4. Their percentage correct will be higher for 'signal' trials in the second half of the experiment