1. (a) By definition of conditional probability, $p(x,y) = p(x|y)p(y) = p(y|x)p(x)$. Thus $p(x) = p(y)p(x|y)/p(y|x)$ and summing both sides over x yields $1 = \sum_x p(x) = p(y)\sum_x p(x|y)p(y|x) \implies p(y) = 1/(\sum_x p(x|y)p(y|x))$.

   (b) $p(x,y) = p(x|y)p(y) = p(x|y)/(\sum_x p(x|y)p(y|x))$ by the previous part, so $p(x,y)$ is determined purely from the conditional distributions $p(x|y)$ and $p(y|x)$. This relates to the Gibbs sampler since it gives some justification that at each step of the Gibb's sampler we can just produce a new sample using the conditional distribution instead of using the joint distribution, since the conditional distributions determine the joint distribution.

2. Define a Markov chain with state space $1, \ldots, N$ labeling the set of all feasible configurations. We define transitions by selecting a random node (uniformly) and then switching whether or not there is a particle there (only adding a particle if it does not create an infeasible configuration). The transitions of this chain are symmetric (since a node is selected uniformly and then switching whether a particle is there or not is done uniformly), so the stationary distribution is uniform over all feasible configurations.

3. (a) Beginning with the pixels in some random configuration (e.g. all black), we cycle through each cell and compute the conditional probability of white and black given the remainder of the grid. This is $P(B = w|Y \setminus B) = P(B = w, Y \setminus B)/P(Y \setminus B) = P(B = w, Y \setminus B)/(P(Y) + P(Y \setminus B \cup B'))$ where the first term of the denominator corresponds to Y and the second term corresponds to Y with the box B flipped and vice versa for $P(B = b|Y|setminus B)$.

   Note that even though the normalizing constant is unknown, it cancels out in the computation here so this does not pose a problem. Furthermore, due to probability being proportional to the exponential of the sum of the indicators (not just the sum), most terms cancel out in the computation and we only need to consider the indicators corresponding to the four adjacent pixels, so each update step is simple to compute only depending on a small sum.

   (b) For a Metropolis Hastings approach, we first construct a Markov chain on the state space of all possible images whose transitions are gotten by uniformly selecting a node and toggling it with probability $1/2$.

   Then accept the transition with probability $P(B_{new}, Y \setminus B)/P(B_{old}, Y \setminus B)$ where B_{new} is the color after flipping and B_{old} is the current color of the pixel we want to flip.