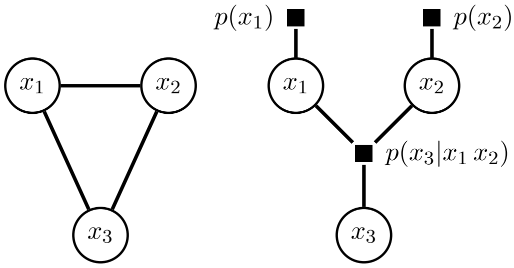
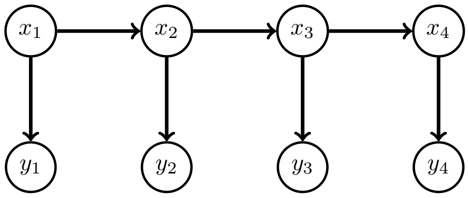
Exercise 1. ***Conversion to factor graphs***

(a) *Draw an undirected graph and an undirected factor graph for* $p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_1, x_2)$

Solution.



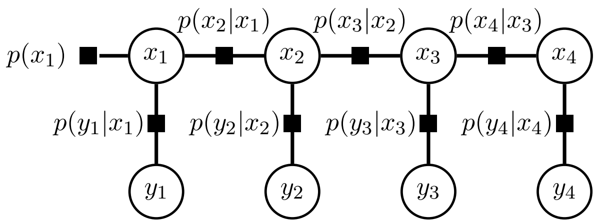
(b) *Draw an undirected factor graph for the directed graphical model defined by the graph below.*



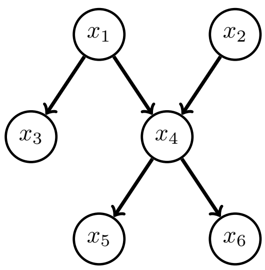
Solution. The graph specifies probabilistic models that factorise as

$$p(x_1, \ldots, x_4, y_1, \ldots, y_4) = p(x_1)p(y_1|x_1)\prod_{i=2}^{4} p(y_i|x_i)p(x_i|x_{i-1})$$

It is the graph for a Hidden Markov model. The corresponding factor graph is shown below.



(c) *Draw the moralised graph and an undirected factor graph for directed graphical models defined by the graph below (this kind of graph is called a polytree: there are no loops but a node may have more than one parent).*

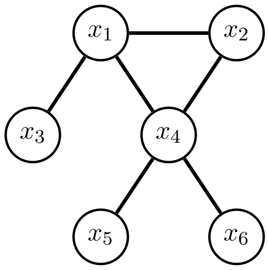


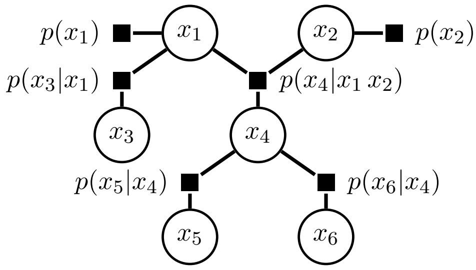
Solution. The moral graph is obtained by connecting the parents of the collider node x_4. See the graph on the left in the figure below.

For the factor graph, we note that the directed graph defines the following class of probabilistic models

$$p(x_1, \ldots x_6) = p(x_1)p(x_2)p(x_3|x_1)p(x_4|x_1, x_2)p(x_5|x_4)p(x_6|x_4)$$

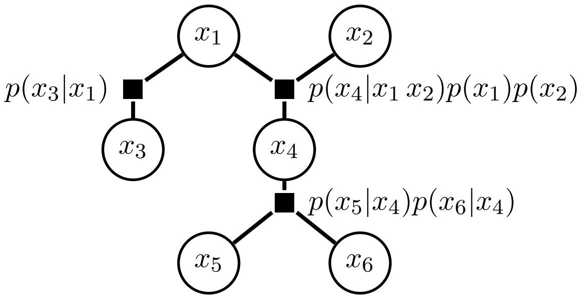
This gives the factor graph on right in the figure below.





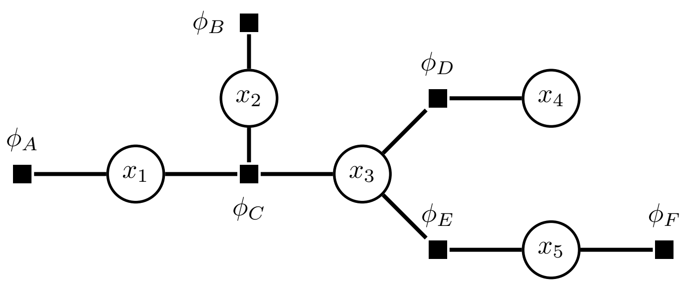
Note:

- The moral graph contains a loop while the factor graph does not. The factor graph is still a polytree. This can be exploited for inference.
- One may choose to group some factors together in order to obtain a factor graph with a particular structure (see factor graph below)



Exercise 2. ***Sum-product message passing***

We here re-consider the factor tree from the lecture on exact inference.



Let all variables be binary, $x_i \in \{0, 1\}$, and the factors be defined as follows:

x_1	ϕ_A
0	2
1	4

x_2	ϕ_B
0	4
1	4

x_1	x_2	x_3	ϕ_C
0	0	0	4
1	0	0	2
0	1	0	2
1	1	0	6
0	0	1	2
1	0	1	6
0	1	1	6
1	1	1	4

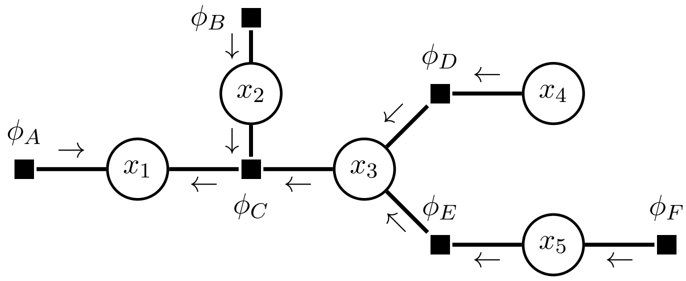
x_3	x_4	ϕ_D
0	0	8
1	0	2
0	1	2
1	1	6

x_3	x_5	ϕ_E
0	0	3
1	0	6
0	1	6
1	1	3

x_5	ϕ_F
0	1
1	8

(a) *Mark the graph with arrows indicating all messages that need to be computed for the computation of* $p(x_1)$*.*

Solution.



(b) *Compute the messages that you have identified.*

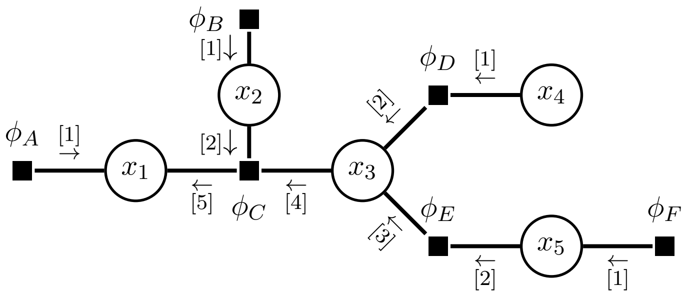
Assuming that the computation of the messages is scheduled according to a common clock, group the messages together so that all messages in the same group can be computed in parallel during a clock cycle.

Solution. Since the variables are binary, each message can be represented as a two-dimensional vector. We use the convention that the first element of the vector corresponds to the message for $x_i = 0$ and the second element to the message for $x_i = 1$. For example,

$$\boldsymbol{\mu_{\phi_A \to x_1}} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} \tag{S.1}$$

means that the message $\mu_{\phi_A \to x_1}(x_1)$ equals 2 for $x_1 = 0$, i.e. $\mu_{\phi_A \to x_1}(0) = 2$.

The following figure shows a grouping (scheduling) of the computation of the messages.



Clock cycle 1:

$$\boldsymbol{\mu_{\phi_A \to x_1}} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} \qquad \boldsymbol{\mu_{\phi_B \to x_2}} = \begin{pmatrix} 4 \\ 4 \end{pmatrix} \qquad \boldsymbol{\mu_{x_4 \to \phi_D}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \boldsymbol{\mu_{\phi_F \to x_5}} = \begin{pmatrix} 1 \\ 8 \end{pmatrix} \tag{S.2}$$

Clock cycle 2:

$$\boldsymbol{\mu_{x_2 \to \phi_C}} = \boldsymbol{\mu_{\phi_B \to x_2}} = \begin{pmatrix} 4 \\ 4 \end{pmatrix} \qquad \boldsymbol{\mu_{x_5 \to \phi_E}} = \boldsymbol{\mu_{\phi_F \to x_5}} = \begin{pmatrix} 1 \\ 8 \end{pmatrix} \tag{S.3}$$

Message $\mu_{\phi_D \to x_3}$ is defined as

$$\mu_{\phi_D \to x_3}(x_3) = \sum_{x_4} \phi_D(x_3, x_4) \mu_{x_4 \to \phi_D}(x_4) \tag{S.4}$$

so that

$$\begin{aligned}
\mu_{\phi_D \to x_3}(0) &= \sum_{x_4=0}^{1} \phi_D(0, x_4) \mu_{x_4 \to \phi_D}(x_4) && \text{(S.5)} \\
&= \phi_D(0,0)\mu_{x_4 \to \phi_D}(0) + \phi_D(0,1)\mu_{x_4 \to \phi_D}(1) && \text{(S.6)} \\
&= 8 \cdot 1 + 2 \cdot 1 && \text{(S.7)} \\
&= 10 && \text{(S.8)} \\
\mu_{\phi_D \to x_3}(1) &= \sum_{x_4=0}^{1} \phi_D(1, x_4) \mu_{x_4 \to \phi_D}(x_4) && \text{(S.9)} \\
&= \phi_D(1,0)\mu_{x_4 \to \phi_D}(0) + \phi_D(1,1)\mu_{x_4 \to \phi_D}(1) && \text{(S.10)} \\
&= 2 \cdot 1 + 6 \cdot 1 && \text{(S.11)} \\
&= 8 && \text{(S.12)}
\end{aligned}$$

and thus

$$\boldsymbol{\mu_{\phi_D \to x_3}} = \begin{pmatrix} 10 \\ 8 \end{pmatrix}. \tag{S.13}$$

The above computations can be written more compactly in matrix notation. Let $\boldsymbol{\phi_D}$ be the matrix that contains the outputs of $\phi_D(x_3, x_4)$

$$\boldsymbol{\phi_D} = \begin{pmatrix} \phi_D(x_3 = 0, x_4 = 0) & \phi_D(x_3 = 0, x_4 = 1) \\ \phi_D(x_3 = 1, x_4 = 0) & \phi_D(x_3 = 1, x_4 = 1) \end{pmatrix} = \begin{pmatrix} 8 & 2 \\ 2 & 6 \end{pmatrix}. \tag{S.14}$$

We can then write $\boldsymbol{\mu_{\phi_D \to x_3}}$ in terms of a matrix vector product,

$$\boldsymbol{\mu_{\phi_D \to x_3}} = \boldsymbol{\phi_D} \boldsymbol{\mu_{x_4 \to \phi_D}}. \tag{S.15}$$

Clock cycle 3:

Representing the factor ϕ_E as matrix $\boldsymbol{\phi_E}$,

$$\boldsymbol{\phi_E} = \begin{pmatrix} \phi_E(x_3 = 0, x_5 = 0) & \phi_E(x_3 = 0, x_5 = 1) \\ \phi_E(x_3 = 1, x_5 = 0) & \phi_E(x_3 = 1, x_5 = 1) \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 3 \end{pmatrix}, \tag{S.16}$$

we can write

$$\mu_{\phi_E \to x_3}(x_3) = \sum_{x_5} \phi_E(x_3, x_5)\mu_{x_5 \to \phi_E}(x_5) \tag{S.17}$$

as a matrix vector product,

$$\boldsymbol{\mu_{\phi_E \to x_3}} = \boldsymbol{\phi_E \mu_{x_5 \to \phi_E}} \tag{S.18}$$

$$= \begin{pmatrix} 3 & 6 \\ 6 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 8 \end{pmatrix} \tag{S.19}$$

$$= \begin{pmatrix} 51 \\ 30 \end{pmatrix}. \tag{S.20}$$

Clock cycle 4:
Variable node x_3 has received all incoming messages, and can thus output $\mu_{x_3 \to \phi_C}$,

$$\mu_{x_3 \to \phi_C}(x_3) = \mu_{\phi_D \to x_3}(x_3)\mu_{\phi_E \to x_3}(x_3). \tag{S.21}$$

Using $\odot$ to denote element-wise multiplication of two vectors, we have

$$\boldsymbol{\mu_{x_3 \to \phi_C}} = \boldsymbol{\mu_{\phi_D \to x_3}} \odot \boldsymbol{\mu_{\phi_E \to x_3}} \tag{S.22}$$

$$= \begin{pmatrix} 10 \\ 8 \end{pmatrix} \odot \begin{pmatrix} 51 \\ 30 \end{pmatrix} \tag{S.23}$$

$$= \begin{pmatrix} 510 \\ 240 \end{pmatrix}. \tag{S.24}$$

Clock cycle 5:
Factor node ϕ_C has received all incoming messages, and can thus output $\mu_{\phi_C \to x_1}$,

$$\mu_{\phi_C \to x_1}(x_1) = \sum_{x_2, x_3} \phi_C(x_1, x_2, x_3)\mu_{x_2 \to \phi_C}(x_2)\mu_{x_3 \to \phi_C}(x_3). \tag{S.25}$$

Writing out the sum for $x_1 = 0$ and $x_1 = 1$ gives

$$\mu_{\phi_C \to x_1}(0) = \sum_{x_2, x_3} \phi_C(0, x_2, x_3)\mu_{x_2 \to \phi_C}(x_2)\mu_{x_3 \to \phi_C}(x_3) \tag{S.26}$$
$$= \phi_C(0, x_2, x_3)\mu_{x_2 \to \phi_C}(x_2)\mu_{x_3 \to \phi_C}(x_3) \mid_{(x_2,x_3)=(0,0)} + \tag{S.27}$$
$$\phi_C(0, x_2, x_3)\mu_{x_2 \to \phi_C}(x_2)\mu_{x_3 \to \phi_C}(x_3) \mid_{(x_2,x_3)=(1,0)} + \tag{S.28}$$
$$\phi_C(0, x_2, x_3)\mu_{x_2 \to \phi_C}(x_2)\mu_{x_3 \to \phi_C}(x_3) \mid_{(x_2,x_3)=(0,1)} + \tag{S.29}$$
$$\phi_C(0, x_2, x_3)\mu_{x_2 \to \phi_C}(x_2)\mu_{x_3 \to \phi_C}(x_3) \mid_{(x_2,x_3)=(1,1)} \tag{S.30}$$
$$= 4 \cdot 4 \cdot 510 + \tag{S.31}$$
$$2 \cdot 4 \cdot 510 + \tag{S.32}$$
$$2 \cdot 4 \cdot 240 + \tag{S.33}$$
$$6 \cdot 4 \cdot 240 \tag{S.34}$$
$$= 19920 \tag{S.35}$$
$$\mu_{\phi_C \to x_1}(1) = \sum_{x_2, x_3} \phi_C(1, x_2, x_3)\mu_{x_2 \to \phi_C}(x_2)\mu_{x_3 \to \phi_C}(x_3) \tag{S.36}$$
$$= \phi_C(1, x_2, x_3)\mu_{x_2 \to \phi_C}(x_2)\mu_{x_3 \to \phi_C}(x_3) \mid_{(x_2,x_3)=(0,0)} + \tag{S.37}$$
$$\phi_C(1, x_2, x_3)\mu_{x_2 \to \phi_C}(x_2)\mu_{x_3 \to \phi_C}(x_3) \mid_{(x_2,x_3)=(1,0)} + \tag{S.38}$$
$$\phi_C(1, x_2, x_3)\mu_{x_2 \to \phi_C}(x_2)\mu_{x_3 \to \phi_C}(x_3) \mid_{(x_2,x_3)=(0,1)} + \tag{S.39}$$
$$\phi_C(1, x_2, x_3)\mu_{x_2 \to \phi_C}(x_2)\mu_{x_3 \to \phi_C}(x_3) \mid_{(x_2,x_3)=(1,1)} \tag{S.40}$$
$$= 2 \cdot 4 \cdot 510 + \tag{S.41}$$
$$6 \cdot 4 \cdot 510 + \tag{S.42}$$
$$6 \cdot 4 \cdot 240 + \tag{S.43}$$
$$4 \cdot 4 \cdot 240 \tag{S.44}$$
$$= 25920 \tag{S.45}$$

and hence

$$\boldsymbol{\mu_{\phi_C \to x_1}} = \begin{pmatrix} 19920 \\ 25920 \end{pmatrix} \tag{S.46}$$

After step 5, variable node x_1 has received all incoming messages and the marginal can be computed.

In addition to the messages needed for computation of $p(x_1)$ one can compute *all* messages in the graph in five clock cycles, see Figure 1. This means that *all* marginals, as well as the joints of those variables sharing a factor node, are available after five clock cycles.

(c) *What is $p(x_1 = 1)$?*

Solution. We compute the marginal $p(x_1)$ as

$$p(x_1) \propto \mu_{\phi_A \to x_1}(x_1)\mu_{\phi_C \to x_1}(x_1) \tag{S.47}$$

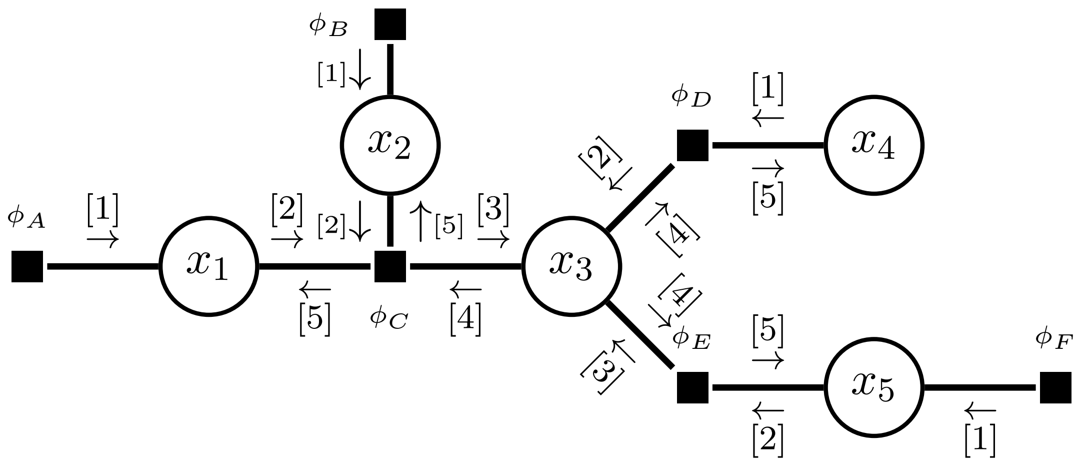


Figure 1: Answer to Exercise 2 Question (b): Computing all messages in five clock cycles. If we also computed the messages toward the leaf factor nodes, we needed six cycles, but they are not necessary for computation of the marginals so they are omitted.

which is in vector notation

$$\begin{pmatrix} p(x_1 = 0) \\ p(x_1 = 1) \end{pmatrix} \propto \boldsymbol{\mu}_{\boldsymbol{\phi_A \to x_1}} \odot \boldsymbol{\mu}_{\boldsymbol{\phi_C \to x_1}} \tag{S.48}$$

$$\propto \begin{pmatrix} 2 \\ 4 \end{pmatrix} \odot \begin{pmatrix} 19920 \\ 25920 \end{pmatrix} \tag{S.49}$$

$$\propto \begin{pmatrix} 39840 \\ 103680 \end{pmatrix}. \tag{S.50}$$

Normalisation gives

$$\begin{pmatrix} p(x_1 = 0) \\ p(x_1 = 1) \end{pmatrix} = \frac{1}{39840 + 103680} \begin{pmatrix} 39840 \\ 103680 \end{pmatrix} \tag{S.51}$$

$$= \begin{pmatrix} 0.2776 \\ 0.7224 \end{pmatrix} \tag{S.52}$$

so that $p(x_1 = 1) = 0.7224$.

Note the relatively large numbers in the messages that we computed. In other cases, one may obtain very small ones depending on the scale of the factors. This can cause numerical issues that can be addressed by working in the logarithmic domain (see Barber's paragraph on log messages, p86 in his book)

(d) *Draw the factor graph corresponding to* $p(x_1, x_3, x_4, x_5 | x_2 = 1)$ *and provide the numerical values for all factors.*

Solution. The pmf represented by the original factor graph is

$$p(x_1, \ldots, x_5) \propto \phi_A(x_1)\phi_B(x_2)\phi_C(x_1, x_2, x_3)\phi_D(x_3, x_4)\phi_E(x_3, x_5)\phi_F(x_5)$$

The conditional $p(x_1, x_3, x_4, x_5|x_2 = 1)$ is proportional to $p(x_1, \ldots, x_5)$ with x_2 fixed to $x_2 = 1$, i.e.

$$p(x_1, x_3, x_4, x_5|x_2 = 1) \propto p(x_1, x_2 = 1, x_3, x_4, x_5) \tag{S.53}$$

$$\propto \phi_A(x_1)\phi_B(x_2 = 1)\phi_C(x_1, x_2 = 1, x_3)\phi_D(x_3, x_4)\phi_E(x_3, x_5)\phi_F(x_5) \tag{S.54}$$

$$\propto \phi_A(x_1)\phi_C^{x_2}(x_1, x_3)\phi_D(x_3, x_4)\phi_E(x_3, x_5)\phi_F(x_5) \tag{S.55}$$

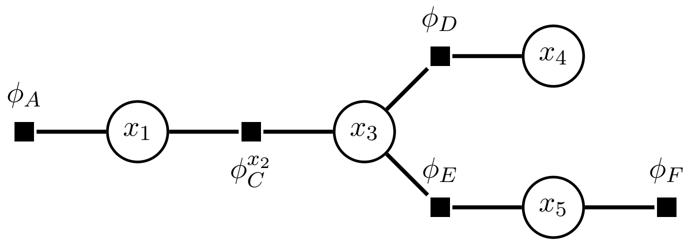
where $\phi_C^{x_2}(x_1, x_3) = \phi_C(x_1, x_2 = 1, x_3)$. The numerical values of $\phi_C^{x_2}(x_1, x_3)$ can be read from the table defining $\phi_C(x_1, x_2, x_3)$, extracting those rows where $x_2 = 1$,

	x_1	x_2	x_3	ϕ_C
	0	0	0	4
	1	0	0	2
→	0	1	0	2
→	1	1	0	6
	0	0	1	2
	1	0	1	6
→	0	1	1	6
→	1	1	1	4

so that

x_1	x_3	$\phi_C^{x_2}$
0	0	2
1	0	6
0	1	6
1	1	4

The factor graph for $p(x_1, x_3, x_4, x_5|x_2 = 1)$ is shown below. Factor ϕ_B has disappeared since it only depended on x_2 and thus became a constant. Factor ϕ_C is replaced by $\phi_C^{x_2}$ defined above. The remaining factors are the same as in the original factor graph.



(e) *Compute $p(x_1 = 1|x_2 = 1)$, re-using messages that you have already computed for the evaluation of $p(x_1 = 1)$.*

Solution. The message $\mu_{\phi_A \to x_1}$ is the same as in the original factor graph and $\mu_{x_3 \to \phi_C^{x_2}} = \mu_{x_3 \to \phi_C}$. This is because the outgoing message from x_3 corresponds to the effective factor obtained by summing out all variables in the sub-trees attached to x_3 (without the $\phi_C^{x_2}$ branch), and these sub-trees do not depend on x_2.

The message $\mu_{\phi_C^{x_2} \to x_1}$ needs to be newly computed. We have

$$\mu_{\phi_C^{x_2} \to x_1}(x_1) = \sum_{x_3} \phi_C^{x_2}(x_1, x_3)\mu_{x_3 \to \phi_C^{x_2}} \tag{S.56}$$

or in vector notation

$$\boldsymbol{\mu}_{\phi_C^{x_2} \to x_1} = \boldsymbol{\phi}_C^{x_2} \boldsymbol{\mu}_{x_3 \to \phi_C^{x_2}} \tag{S.57}$$

$$= \begin{pmatrix} \phi_C^{x_2}(x_1 = 0, x_3 = 0) & \phi_C^{x_2}(x_1 = 0, x_3 = 1) \\ \phi_C^{x_2}(x_1 = 1, x_3 = 0) & \phi_C^{x_2}(x_1 = 1, x_3 = 1) \end{pmatrix} \boldsymbol{\mu}_{x_3 \to \phi_C^{x_2}} \tag{S.58}$$

$$= \begin{pmatrix} 2 & 6 \\ 6 & 4 \end{pmatrix} \begin{pmatrix} 510 \\ 240 \end{pmatrix} \tag{S.59}$$

$$= \begin{pmatrix} 2460 \\ 4020 \end{pmatrix} \tag{S.60}$$

We thus obtain for the marginal posterior of x_1 given $x_2 = 1$:

$$\begin{pmatrix} p(x_1 = 0 | x_2 = 1) \\ p(x_1 = 1 | x_2 = 1) \end{pmatrix} \propto \boldsymbol{\mu}_{\phi_A \to x_1} \odot \boldsymbol{\mu}_{\phi_C^{x_2} \to x_1} \tag{S.61}$$

$$\propto \begin{pmatrix} 2 \\ 4 \end{pmatrix} \odot \begin{pmatrix} 2460 \\ 4020 \end{pmatrix} \tag{S.62}$$

$$\propto \begin{pmatrix} 4920 \\ 16080 \end{pmatrix}. \tag{S.63}$$

Normalisation gives

$$\begin{pmatrix} p(x_1 = 0 | x_2 = 1) \\ p(x_1 = 1 | x_2 = 1) \end{pmatrix} = \begin{pmatrix} 0.2343 \\ 0.7657 \end{pmatrix} \tag{S.64}$$

and thus $p(x_1 = 1 | x_2 = 1) = 0.7657$. The posterior probability is slightly larger than the prior probability, $p(x_1 = 1) = 0.7224$.