14.6 Let H_x be a random variable denoting the handedness of an individual x, with possible values l or r. A common hypothesis is that left- or right-handedness is inherited by a simple mechanism; that is, perhaps there is a gene G_x, also with values l or r, and perhaps actual handedness turns out mostly the same (with some probability s) as the gene an individual possesses. Furthermore, perhaps the gene itself is equally likely to be inherited from either of an individual's parents, with a small nonzero probability m of a random mutation flipping the handedness.

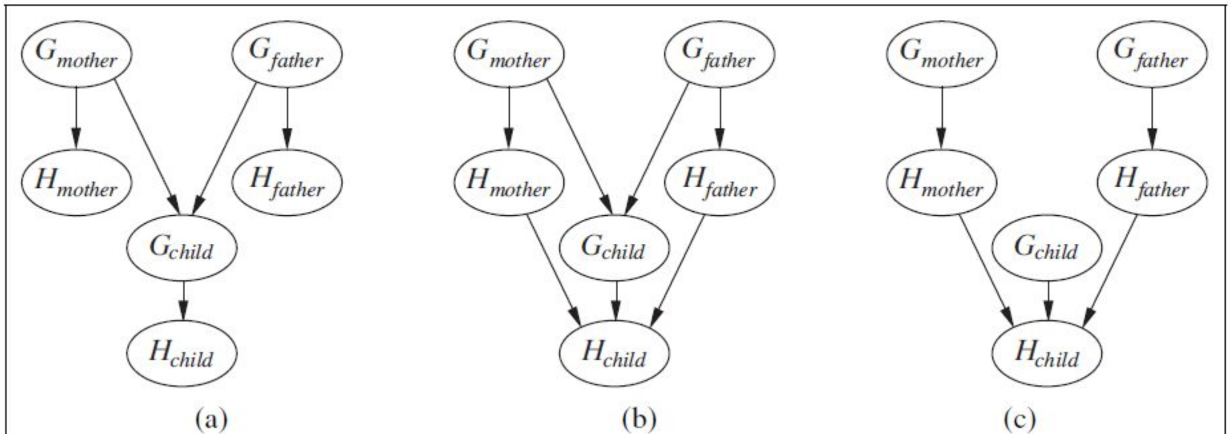

Figure 14.20 Three possible structures for a Bayesian network describing genetic inheritance of handedness.

a. Which of the three networks in Figure 14.20 claim that:
$P(G_{father}.G_{mother}, G_{child}) = P(G_{father}).P(G_{mother}).P(G_{child})$ **?**

The above formula says that the three variables G_x are independent. The Bayesian Network (c) is the only one that presents a topological order where these variables are independent.

We can check this using equation 14.1 as a reference for the 3 Bayesian networks:
$$P(x_1,\ x_2,\ \ldots,\ x_n) = \prod_{i=1}^{n} \theta(x_i \mid parents(X_i))$$

$(a)\ P(G_f, G_m, G_c, H_f, H_m, H_c) = P(G_f).P(G_m).P(G_c|\ G_m, G_f).P(H_c|\ G_c).P(H_m|G_m).P(H_f|G_f)$
$(a)\ P(G_f, G_m, G_c) = P(G_f).P(G_m).P(G_c|\ G_m, G_f)$

$(b)\ P(G_f, G_m, G_c, H_f, H_m, H_c) = P(G_m).P(G_f).P(G_c|\ G_m, G_f).P(H_c|\ G_c, H_m, H_f).P(H_m|G_m).P(H_f|G_f)$
$(b)\ P(G_f, G_m, G_c) = P(G_f).P(G_m).P(G_c|\ G_m, G_f)$

$(c)\ P(G_f, G_m, G_c, H_f, H_m, H_c) = P(G_f).P(G_m).P(G_c).P(H_c|\ G_c, H_m, H_f).P(H_m|G_m).P(H_f|G_f)$
$(c)\ P(G_f, G_m, G_c) = P(G_f).P(G_m).P(G_c)$

We can see above that only network (c) meets independence definition for $P(G_f, G_m, G_c)$

b. Which of the three networks make independence claims that are consistent with the hypothesis about the inheritance of handedness?

Network (a) claims that H_c is independent of H_f and H_m and this is consistent with the hypothesis.
Network (b) claims that H_c depends of G_c and G_c depends of its parents and it is consistent too.
Network (c) is not consistent with the hypothesis because G_c does not depend on its parents.

Answer: Networks (a) and (b)

c. Which of the three networks is the best description of the hypothesis?

Network (a) is the best description because Network (b) says that H_c depends on H_f and H_m and this statement is not mentioned in the hypothesis and in some way it add more noise to the network. But for example network (b) could be used if in our hypothesis we suspect that observation of parent's handedness can influence child's handedness (child is the observer).

d. Write down the CPT for the G_{child} node in network (a), in terms of s and m.

G_{child}	G_{father}	G_{mother}	$P(G_c \| G_f,\ G_m)$
$left$	$right$	$right$	m
$left$	$right$	$left$	1/2
$left$	$left$	$right$	1/2
$left$	$left$	$left$	$1 - m$

If variable $G_{child} = right$ we can use the complement probability with $G_{child} = left$ to resolve it. For example:

$$P(G_c = r \mid G_f = r,\ G_m = r) = 1 - P(G_c = l \mid G_f = r,\ G_m = r) = 1 - m$$

e. Suppose that $P(G_{father} = l) = P(G_{mother} = l) = q$. In network (a), derive an expression for $P(G_{child} = l)$ in terms of m and q only, by conditioning on its parent nodes.

By marginalization rule:

$$P(G_{child} = l) = \sum_{x \in G_f, y \in G_m} P(G_c = l,\ G_f = x, G_m = y)$$

By product rule:

$$P(G_{child} = l) = \sum_{x \in G_f, y \in G_m} P(G_c = l \mid G_f = x, G_m = y) \times P(G_f = x, G_m = y)$$

By independence rule:

$$P(G_{child} = l) = \sum_{x \in G_f, y \in G_m} P(G_c = l \mid G_f = x, G_m = y) \times P(G_f = x) \times P(G_m = y)$$

$$
\begin{aligned}
&P(G_{child} = l) = \\
&+ P(G_c = l \mid G_f = l, G_m = l) \times P(G_f = l) \times P(G_m = l) \\
&+ P(G_c = l \mid G_f = r, G_m = l) \times P(G_f = r) \times P(G_m = l) \\
&+ P(G_c = l \mid G_f = l, G_m = r) \times P(G_f = l) \times P(G_m = r) \\
&+ P(G_c = l \mid G_f = r, G_m = r) \times P(G_f = r) \times P(G_m = r)
\end{aligned}
$$

Now we can replace with values calculated in item d) and values defined in item e):

$$
\begin{aligned}
&P(G_{child} = l) = [(1-m) \times q \times q] + [(1/2) \times (1-q) \times q] + [(1/2) \times q \times (1-q)] + [m \times (1-q) \times (1-q)] \\
&P(G_{child} = l) = q^2 - q^2 m + (q/2) - (q^2/2) + (q/2) - (q^2/2) + m - 2mq \ + mq^2 \\
&P(G_{child} = l) = q^2 - q^2 m + q - q^2 + m - 2mq \ + mq^2 \\
&P(G_{child} = l) = q + m - 2mq
\end{aligned}
$$

Then we can calculate $P(G_{child} = l)$ based on values of q and m .

f. Under conditions of genetic equilibrium, we expect the distribution of genes to be the same across generations. Use this to calculate the value of q , and, given what you know about handedness in humans, explain why the hypothesis described at the beginning of this question must be wrong.

Given that $P(G_{father} = l) = P(G_{mother} = l) = q$ if there exists genetic equilibrium then $P(G_{child} = l) = q$

And using the formula we already found in item e) we can say that:
$P(G_{child} = l) = q + m - 2mq = q \Leftrightarrow m - 2mq = 0 \Leftrightarrow m(1-2q) = 0$
$m(1-2q) = 0 \Leftrightarrow m = 0 \vee 1 - 2q = 0 \Leftrightarrow m = 0 \vee q = 1/2$

If $m = 0$ then there exists no mutation, that is not consistent with hypothesis.
So, the only possible is that value of $q = 1/2$.

But then:
$P(G_{father} = l) = P(G_{mother} = l) = P(G_{child} = l) = 1/2$

And then:
$P(G_{father} = r) = P(G_{mother} = r) = P(G_{child} = r) = 1/2$

And we know by simple experience that the probability of being right-handedness is bigger than left-handedness. Then the handedness can not be supported only on parents genes. So the hypothesis must be wrong.

Q.E.D.

Handedness displays a complex inheritance pattern. For example, if both parents of a child are left-handed, there is a 26% chance of that child being left-handed [https://en.wikipedia.org/wiki/Handedness]

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=handedness+hereditary+genetics