Q1. Refer to Question 1 section in the R script for code.

E1 posterior probability using my priors: 0.1238938

E2 posterior probability with updated priors: 0.6644068

E1 posterior probability using my colleague's priors: 0.9962547

E2 posterior probability with updated priors: 0.9997315

As shown in this example, the model produces very different estimates of the probability of a true difference existing depending on the priors used. When I calculated the posterior probability using my prior (dis)beliefs regarding the intelligence boosting properties of red M&Ms, the posterior probability of a true difference existing when a significant difference was found was only around 12%. On the other hand, using my colleague's priors (i.e., strong beliefs that M&Ms increase intelligence) would lead me to the conclusion that there is a >99% chance a true difference exists given that one was observed. This can be seen as analogous to controversial disagreements such as the existence of ghosts. For example, a ghost hunter might have strong prior beliefs in the existence of ghosts (e.g., there is a 99% chance that ghosts exist). If this person were to conduct an experiment with a low false positive rate (e.g., 5%) and observe evidence for the existence of ghosts, then, these strong prior beliefs might lead them to believe that there is a high probability that ghosts exist given that evidence was observed. On the other hand, a skeptic who is unconvinced that ghosts exist may have prior beliefs that allow only a very small probability that ghosts exist (e.g., there is a 1% chance that ghosts exist). If the skeptic considered the ghost hunter's experiment, these priors would lead the skeptic to believe that even given evidence for the existence of ghosts and the experiment's low false positive rate, it is still highly unlikely that the entities truly exist. Despite observing the same evidence, the skeptic and the ghost hunter would arrive at very different conclusions due to their prior beliefs.

Q2. Refer to Question 2 section in the R script for code.

Frequentist t-test:

```
Paired t-test

data:  extra by group
t = -4.0621, df = 9, p-value = 0.002833
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -2.4598858 -0.7001142
sample estimates:
mean difference
          -1.58
```

Bayes factor t-test:

```
Bayes factor analysis
--------------
[1] Alt., r=0.707 : 17.25888 ±0%

Against denominator:
  Null, mu = 0
---
Bayes factor type: BFoneSample, JZS
```

For the *sleep* dataset, H_0 represent the null hypothesis: There is no true difference in the number of extra hours of sleep had between the two drugs administered in the population. H_1 represents the alternative hypothesis: There is a true difference in the number of extra hours of sleep had between the two drugs administered in the population. The Frequentist *p*-value of .003 tells us that there is a ~0.3% chance of observing a result at least as extreme as the difference observed given that the null hypothesis is true; were the experiment repeated many times and the null hypothesis true, we would expect to observe a result this extreme 0.3% of the time. Accordingly, the *p*-value tells us that it is extremely unlikely we would observe the relationship we did if the null hypothesis was true. According to conventional standards for statistical significance in psychology, this value tells us that the null hypothesis should be rejected. Given that the *p*-value is based on the assumption that the null hypothesis is true, however, it does not directly provide support for the alternative hypothesis.

On the other hand, the Bayesian t-test provides clear support for the alternative hypothesis (i.e., that the difference between conditions in the population is not equal to zero) over the null hypothesis (i.e., that the difference between conditions in the population is equal to zero). This test produced a BF_{10} of 17.26, which tells us that the alternative model is 17.26 times more likely than the null model. Similarly, inverting this Bayes factor produces a BF_{01} of 0.06. This indicates that the null model is 0.06 times more likely than the alternative model. As such, the alternative model is clearly favored here.

Q3. Refer to Question 3 section in the R script for code.

Frequentist t-test:

```
Welch Two Sample t-test

data:  len by supp
t = 1.9153, df = 55.309, p-value = 0.06063
alternative hypothesis: true difference in means between group OJ and group
VC is not equal to 0
95 percent confidence interval:
 -0.1710156  7.5710156
sample estimates:
mean in group OJ mean in group VC
        20.66333         16.96333
```

Bayes factor t-test:

```
Bayes factor analysis
--------------
[1] Alt., r=0.707 : 1.198757 ±0.01%

Against denominator:
  Null, mu1-mu2 = 0
---
Bayes factor type: BFindepSample, JZS
```

For this analysis of the *toothgrowth* dataset, H_0 represent the null hypothesis: There is no true difference in tooth length between methods of vitamin C administration in the population. H_1

represents the alternative hypothesis: There is no true difference in the population in tooth length between methods of vitamin C administration in the population. The Frequentist *p*-value of .061 tells us only that there is a ~6% chance of observing a result at least as extreme as the difference observed given that the null hypothesis is true; were the experiment repeated many times and the null hypothesis true, we would expect to observe a result this extreme 6% of the time. Accordingly, the *p*-value tells us that it is unlikely – but not extremely unlikely – that we would observe the relationship we did if the null hypothesis was true. According to conventional standards for statistical significance in psychology, this value tells us that the null hypothesis cannot be rejected. Given that the *p*-value is based on the assumption that the null hypothesis is true, however, it does not directly tell us anything about the alternative hypothesis.

The Bayesian t-test provides weak support for the alternative hypothesis (i.e., that the difference between conditions in the population is not equal to zero) over the null hypothesis (i.e., that the difference between conditions in the population is equal to zero). This test produced a BF_{10} of 1.20, which tells us that the alternative model is 1.20 times more likely than the null model. Similarly, inverting this Bayes factor produces a BF_{01} of 0.83. This indicates that the null model is 0.83 times more likely than the alternative model. As such, the alternative model is favored by Bayesian evidence here, but only weakly.

Q4. Refer to Question 4 section in the R script for code.

Frequentist t-test:

```
Welch Two Sample t-test

data:  len by supp
t = 1.9153, df = 55.309, p-value = 0.9697
alternative hypothesis: true difference in means between group OJ and group
VC is less than 0
95 percent confidence interval:
     -Inf 6.931731
sample estimates:
mean in group OJ mean in group VC
```

```
        20.66333         16.96333
```

Bayesian t-test:

```
Bayes factor analysis
--------------
[1] Alt., r=0.707 -Inf<d<0    : 0.1001702 ±0.04%
[2] Alt., r=0.707 !(-Inf<d<0) : 2.297343  ±0%

Against denominator:
  Null, mu1-mu2 = 0 
---
Bayes factor type: BFindepSample, JZS
```

For this analysis of the *toothgrowth* dataset, H_0 represent the null hypothesis: The true mean difference in tooth length between the orange juice and vitamin C groups is not greater than zero in the population. H_1 represents the alternative hypothesis: The true mean difference in tooth length between the orange juice and vitamin C groups is greater than zero in the population. The Frequentist *p*-value of .970 tells us that there is a ~97% chance of observing a result at least as extreme as the difference observed given that the null hypothesis is true; were the experiment repeated many times and the null hypothesis true, we would expect to observe a result this extreme 97% of the time. Accordingly, the *p*-value tells us that it is not unlikely that we would observe the relationship we did if the null hypothesis was true. According to conventional standards for statistical significance in psychology, this value tells us that the null hypothesis cannot be rejected. Given that the *p*-value is based on the assumption that the null hypothesis is true, however, it does not inform us about the alternative hypothesis.

The Bayesian t-test does not support the alternative hypothesis (i.e., that the mean difference in the population is less than zero) over the null hypothesis (i.e., that the difference between the conditions in the population is equal to zero). This test produced a BF_{10} of 0.10, which tells us that the alternative model is 0.10 times more likely than the null model. Similarly, inverting this Bayes factor produces a BF_{01} of 9.98. This indicates that the null model is 9.98 times more likely than the alternative model. As such, the null model is favored by Bayesian evidence here.

Q5. Refer to Question 5 section in R script for code.

Frequentist ANOVA

```
$ANOVA
     Effect DFn DFd      SSn     SSd         F            p p<.05       ges
1      supp   1  54  205.350 712.106 15.571979 2.311828e-04     * 0.2238254
2      dose   2  54 2426.434 712.106 91.999965 4.046291e-18     * 0.7731092
3 supp:dose   2  54  108.319 712.106  4.106991 2.186027e-02     * 0.1320279

$`Levene's Test for Homogeneity of Variance`
  DFn DFd    SSn     SSd        F         p p<.05
1   5  54 38.926 246.053 1.708578 0.1483606
```

Bayesian ANOVA

```
Bayes factor analysis
--------------
[1] supp                     : 1.198757     ±0.01%
[2] dose                     : 4.983636e+12 ±0%
[3] supp + dose              : 2.88581e+14  ±1.59%
[4] supp + dose + supp:dose : 7.626021e+14 ±1.38%

Against denominator:
  Intercept only
---
Bayes factor type: BFlinearModel, JZS
```

Interaction vs both effects:

$BF_{10} = 2.77$

Supp vs both effects:

$BF_{10} = 4.35 \times 10^{-15}$

Dose vs both effects:

$BF_{10} = 0.02$

The Frequentist *p*-value of $p < .001$ for the effect of supp tells us that we have less than 0.1% probability of observing a result at least as extreme as the observed result assuming the null hypothesis is true (i.e., there is no effect of supp). Conventional thresholds for significance suggest that because this result is extremely unlikely, the null hypothesis should be rejected.

The Frequentist *p*-value of $p < .001$ for the effect of dose tells us that we have less than 0.1% probability of observing a result at least as extreme as the observed result assuming the null hypothesis is true (i.e., there is no effect of dose). Conventional thresholds for significance suggest that because this result is extremely unlikely, the null hypothesis should be rejected.

The Frequentist *p*-value of $p = .022$ for the interaction between supp and dose tells us that we have a 2.2% probability of observing a result at least as extreme as the observed result assuming the null hypothesis is true (i.e., there is no effect of the interaction between supp and dose). Conventional thresholds for significance suggest that because this result is extremely unlikely, the null hypothesis should be rejected.

The BF_{10} of 2.77 for the comparison between the omnibus model (i.e., the model including the effect of dose, supp and the interaction between the two main effects) and the main effects-only model supports inclusion of the interaction. This Bayes factor tells us that the omnibus model is 2.77 times more likely than the model including only main effects. Inverting this Bayes factor indicates that the main effects-only model is 0.36 times more likely than the omnibus model. Accordingly, the omnibus model is favored by Bayesian evidence; it appears that accounting for the interaction better explains the data than accounting only for main effects, although the relatively small Bayes factor indicates that this conclusion is not necessarily decisive.

The BF_{10} of 4.35×10^{-15} for the comparison between the model including only supp and the model including both main effects supports the latter. This Bayes factor tells us that the supp-only model is 4.35×10^{-15} times more likely than the model including both main effects. Inverting this Bayes factor indicates that the two-effect model is 2.30×10^{14} times more likely than the supp-only model. Accordingly, the two-effect model is overwhelmingly favored by

Bayesian evidence; it appears that accounting for both effects better explains the data than accounting only for the effect of supp.

The BF$_{10}$ of 0.02 for the comparison between the model including only dose and the model including both main effects supports the latter. This Bayes factor tells us that the dose-only model is 0.02 times more likely than the model including both main effects. Inverting this Bayes factor indicates that the two-effect model is 55.34 times more likely than the dose-only model. Accordingly, the two-effect model is strongly favored by Bayesian evidence; it appears that accounting for both effects better explains the data than accounting only for the effect of dose.

Q6. Refer to Question 6 section in the R script for code.

```
Inclusion Bayes Factors (Model Averaged)

          P(prior) P(posterior) Inclusion BF
supp          0.60         1.00       139.40
dose          0.60         1.00     3.17e+14
dose:supp     0.20         0.73        10.94

* Compared among: all models
*    Priors odds: uniform-equal
```

The inclusion BF of 139.40 for the effect of supp indicates that models including the effect of supp are 139.40 times more likely than models that do not include the effect of supp. This represents strong evidence for the effect of supp.

The inclusion BF of 3.17×10^{14} for the effect of dose indicates that models including the effect of dose are 3.17×10^{14} times more likely than models that do not include the effect of dose. This represents strong evidence for the effect of dose.

The inclusion BF of 10.94 for the interaction between supp and dose indicates that models including the interaction term are 10.94 times more likely than models that do not include the interaction term. This represents substantial evidence for the interaction between dose and supp.

Q7. Refer to Question 7 section of the R script for code.

Medium preset:

```
Bayes factor analysis
--------------
[1] Alt., r=0.707 : 1.198757 ±0.01%

Against denominator:
  Null, mu1-mu2 = 0
---
Bayes factor type: BFindepSample, JZS
```

Wide preset:

```
Bayes factor analysis
--------------
[1] Alt., r=1 : 0.992608 ±0.02%

Against denominator:
  Null, mu1-mu2 = 0
---
Bayes factor type: BFindepSample, JZS
```

Ultra wide preset:

```
Bayes factor analysis
--------------
[1] Alt., r=1.414 : 0.7780345 ±0.03%

Against denominator:
  Null, mu1-mu2 = 0
---
Bayes factor type: BFindepSample, JZS
```

The output differs depending on the prior preset used: As the prior widens, evidence increasingly fails to support the alternative model (i.e., a difference between groups). Moving from narrow to wide priors increases the standard deviation of the prior distribution and accordingly allocates additional probability to larger effects (either positive or negative, assuming a two-tailed test). In the case of this example, the difference between the VC and OJ groups is relatively small. Accordingly, allocating additional probability to larger effects decreases the sensitivity of the analysis to small effects (relative to narrow priors), such as the effect that is potentially present in this example; this is reflected in the decrease in confidence that the alternative model is more likely than the null. Using wide or ultrawide priors may be defensible in scenarios wherein the researcher has absolutely no idea what effect is expected. In such cases, one may want to allocate probability more evenly across potential effects.

Q8. Refer to Question 8 section in R script for code.

Coin 1

Probability that coin is fair: 0.1882621

Probability that coin is heads-biased: 0.8085794

Probability that coin is tails-biased: 0.003158513

This coin is most likely heads-biased.

Coin 2

Probability that coin is fair: 0.4327655

Probability that coin is heads-biased: 0.1134469

Probability that coin is tails-biased: 0.4537876

This coin is most likely tails-biased.

Coin 3

Probability that coin is fair: 0.328084

Probability that coin is heads-biased: 0.5375328

Probability that coin is tails-biased: 0.1343832

This coin is most likely heads-biased.

Coin 4

Probability that coin is fair: 0.1282051

Probability that coin is heads-biased: 0.8205128

Probability that coin is tails-biased: 0.05128205

This coin is most likely heads-biased.

Q9. Refer to Question 9 section in R script for code.

Coin 1

Probability that coin is fair: 0.1399319

Probability that coin is 80% heads: 0.601003

Probability that coin is 80% tails: 0.002347668

Probability that coin is all heads: 0

Probability that coin is 60% heads: 0.2567175

This coin is most likely the 80% heads coin.

Coin 2

Probability that coin is fair: 0.3312894

Probability that coin is 80% heads: 0.08684553

Probability that coin is 80% tails: 0.3473821

Probability that coin is all heads: 0

Probability that coin is 60% heads: 0.2344829

This coin is most likely the 80% tails coin.

Coin 3

Probability that coin is fair: 0.2407365

Probability that coin is 80% heads: 0.3944226

Probability that coin is 80% tails: 0.09860565

Probability that coin is all heads: 0

Probability that coin is 60% heads: 0.2662353

This coin is most likely the 80% heads coin.

Coin 4

Probability that coin is fair: 0.03961965

Probability that coin is 80% heads: 0.2535658

Probability that coin is 80% tails: 0.01584786

Probability that coin is all heads: 0.6339144

Probability that coin is 60% heads: 0.0570523

This coin is most likely the all heads coin.

Q10. My posterior belief that a given coin will always turn up heads often drops off very quickly because the data were – in most cases – wholly incompatible with that belief. Because the probability of obtaining heads for an all-heads coin is 100%, even a singular observation of tails informs the model that the coin being flipped cannot possibly be the all-heads coin. It would not matter if we had a pile of 10000 coins, 9999 of which were all-heads coins and one of which was fair; as soon as an incompatible observation arises, my posterior belief that the coin I picked is all-heads will immediately drop to zero. However, this does not occur for Coin 4 in Question 9. This is simply because Coin 4 produced only heads, thereby providing no observations that were incompatible with an all-heads coin.