Appendix

Sally is deciding which religion to believe. Let W be the set of possible worlds, the sample space on which we'll be assigning subjective probabilities. Decide on a finite partition of W into n sets, say,

$$W = R_1 \cup R_2 \cup \ldots \cup R_n$$

where R_1 is the set of possible worlds in which the first religion we're considering (say, atheism) is true, R_2 is the set of possible worlds in which the second (say, Baha'i) is true, ..., and R_n is the set of possible worlds in which the other religions we haven't considered yet are true.

Let $F \subset W$ be the worlds in which Sally is freely able to choose to place faith in a particular religion; $F^c := W - F$ is the set of worlds in which she is predestined or randomly forced to adopt the religion that she does.

If Sally has free will, let a be an action that Sally could decide to adopt toward religion. a might represent believing a particular form of Christianity, or it might represent becoming a staunch atheist and encouraging others to do the same. She must take *some* action, though, even if it is only to become agnostic.

Let $H \subset W$ be the possible worlds in which Sally experiences absolute infinite suffering (presumably, though not necessarily, in an eternal hell). If Sally has free will and chooses action a, let $\mathbb{P}_a(H|F)$ represent Sally's probability of enduring absolute infinite suffering as a result. Let $\mathbb{P}_a(H)$ represent Sally's overall probability of enduring absolutely infinite suffering, whether after freely taking action a if she has free will or doing what she was predetermined or randomly forced to do if she doesn't have free will. Sally's goal is to choose the action a that will minimize $\mathbb{P}_a(H)$.[1]

By definition of conditional probability,

$$\mathbb{P}_a(H) = \mathbb{P}_a(H|F)\mathbb{P}(F) + \mathbb{P}(H|F^c)\,\mathbb{P}(F^c)\,.$$

Note that $\mathbb{P}(H|F^c)$ is the same for all a, because if Sally has no free will in the area of salvation, then she can't freely choose the action that she will take toward religion. Thus, if Sally wants to minimize $\mathbb{P}_a(H)$, it suffices to minimize $\mathbb{P}_a(H|F)\mathbb{P}(F)$, because the term $\mathbb{P}_a(H|F^c)\,\mathbb{P}(F^c)$ is constant for all a.[2] In fact, since $\mathbb{P}(F)$ is constant too, it suffices to minimize $\mathbb{P}_a(H|F)$.

Again by definition of conditional probability,

$$\mathbb{P}_a(H|F) = \sum_{i=1}^{n} \mathbb{P}_a(H|F \cap R_i)\,\mathbb{P}(R_i|F)\,.$$

According to Bayes' Theorem, $\mathbb{P}(R_i|F) = \mathbb{P}(F|R_i)\frac{\mathbb{P}(R_i)}{\mathbb{P}(F)}$. As I noted in the introduction to this piece, I have assumed for the sake of fairness that $\mathbb{P}(R_i)$ is the same for all i. Thus, $\frac{\mathbb{P}(R_i)}{\mathbb{P}(F)}$ is constant with respect to i, and we have

$$\mathbb{P}_a(H|F) = \frac{\mathbb{P}(R_i)}{\mathbb{P}(F)} \sum_{i=1}^{n} \mathbb{P}_a(H|F \cap R_i)\,\mathbb{P}(F|R_i)\,.$$

[1] There are other possibilities. If $J \subset W$ represents the possible worlds in which Sally experiences absolute infinite joy, then she might instead aim to maximize $\mathbb{P}_a(J)$. Or, more generally, she might try to maximize $\mathbb{P}_a(J) - \lambda\mathbb{P}_a(H)$ for some value of λ.

[2] See "Why We Should Believe in Free Will," (http://utilitarian-essays.com/free-will.html) for more on this.

It suffices to minimize this quantity:

$$\sum_{i=1}^{n} \mathbb{P}_a\left(H|F \cap R_i\right) \mathbb{P}\left(F|R_i\right).$$

Let $E \subset W$ be the worlds in which hell (or, in the case of atheism, some sort of suffering) is eternal. Then we want to minimize

$$\sum_{i=1}^{n} \left[\mathbb{P}\left(H|F \cap R_i \cap E\right) \mathbb{P}\left(E|F \cap R_i\right) + \mathbb{P}_a\left(H|F \cap R_i \cap E^c\right) \mathbb{P}\left(E^c|F \cap R_i\right)\right] \mathbb{P}\left(F|R_i\right).$$

I'll assume $\mathbb{P}_a\left(H|F \cap R_i \cap E^c\right) = 0$ (a temporary hell is not absolutely infinitely bad) so that we have

$$\sum_{i=1}^{n} \mathbb{P}_a\left(H|F \cap R_i \cap E\right) \mathbb{P}\left(E|F \cap R_i\right) \mathbb{P}\left(F|R_i\right).$$

Let $A \subset W$ be the worlds in which hell is not only eternal but absolutely infinitely bad. Now we want to minimize

$$\sum_{i=1}^{n} \left[\mathbb{P}_a\left(H|F \cap R_i \cap E \cap A\right) \mathbb{P}\left(A|F \cap R_i \cap E\right) + \mathbb{P}_a\left(H|F \cap R_i \cap E \cap A^c\right) \mathbb{P}\left(A^c|F \cap R_i \cap E\right)\right] \mathbb{P}\left(E|F \cap R_i\right) \mathbb{P}\left(F|R_i\right).$$

By definition, $\mathbb{P}_a\left(H|F \cap R_i \cap E \cap A^c\right) = 0$, which leaves

$$\sum_{i=1}^{n} \mathbb{P}_a\left(H|F \cap R_i \cap E \cap A\right) \mathbb{P}\left(A|F \cap R_i \cap E\right) \mathbb{P}\left(E|F \cap R_i\right) \mathbb{P}\left(F|R_i\right).$$

In this piece, I've assumed that $\mathbb{P}\left(A|F \cap R_i \cap E\right)$ is the same for all i. Hence, it pulls out of the summation, wherefore we need only minimize

$$\boxed{\sum_{i=1}^{n} \mathbb{P}_a\left(H|F \cap R_i \cap E \cap A\right) \mathbb{P}\left(E|F \cap R_i\right) \mathbb{P}\left(F|R_i\right).}$$

In the body of the essay, I have usually just commented on $\mathbb{P}\left(E|R_i\right)$ and $\mathbb{P}_a\left(H|R_i\right)$, assuming that they are fairly close to $\mathbb{P}\left(E|F \cap R_i\right)$ and $\mathbb{P}_a\left(H|F \cap R_i \cap E \cap A\right)$, respectively. However, this simplification isn't necessary.

I leave it as an exercise to the reader to see how considerations 2-5 in the body of the essay may be derived from the boxed expression.