Support the author!
If you like this book, buy a copy!

Appendix A

Purely Quantitative & Logic Answers

This appendix contains answers to the questions posed in chapter 1.

Answer 1.1: This question has appeared over and over again. Although simple, it is rarely answered well. No calculation is required to determine the answer. If you used *any* algebra whatsoever, stop now, go back, reread the question, and try again.

When the quantity Q of water is poured into the alcohol jug, the concentration of alcohol in the alcohol jug becomes $\frac{V}{V+Q}$. After mixing and pouring some back, the concentration of alcohol in the alcohol jug does not change again (because no new water is added). However, when the diluted alcohol is poured back into the water jug, the concentration of water in the water jug changes from 100% to $\frac{V}{V+Q}$. That is, the final concentrations are identical.

How do you see that the final concentrations must be identical? Remember, you do not need any calculations at all. In fact, the only reason for any calculation is if you also want to find out what the final concentrations are (you were not asked this, but if you wish to work it out, your calculations need not go beyond those of the previous paragraph).

Here is how it works. At the end of the process, both jugs contain the same volume of fluid as they did at the start. The only way for the concentration of alcohol (for example) to have changed from 100% is if some alcohol was displaced by water. Similarly, the only way for the concentration of water to have changed from 100% is if some water was displaced by alcohol. Volume is conserved (both total volume and volume in each jug), so all that has happened is that identical quantities of water and alcohol have traded places (and these identical quantities are slightly less than Q). By symmetry, the concentrations of alcohol in the alcohol jug and water in the water jug must be identical.

If you are still stuck, here is another way of thinking about it. Imagine a black bucket with 1,000 black marbles in it and a white bucket with 1,000 white marbles in it. Suppose I take 100 black marbles out of the black bucket and put them in the white bucket and mix it up really well. Then I have 1,100 marbles in the white bucket, and the great majority of them are white. Suppose I then take 100 marbles from those 1,100 and use them to top the black bucket back up to 1,000 marbles. Then both buckets have 1,000 marbles again. Let us suppose that 91 (i.e., the great majority) of the 100 marbles used to top up the black bucket were white. That means I must have returned only 9 of the original 100 black marbles back to the black bucket when I topped it up. That means I must have left 91 black marbles behind in the white bucket—the same as the number of white marbles that migrated over to the black bucket. So, the proportions are identical!

Answer 1.2: This is a very common question, and a very simple one. You need to figure out the sum: $1+2+3+\cdots+99+100$. There are several ways to do this.

FIRST SOLUTION

A simple technique is to note that the first and last terms add to 101. The second and second-to-last terms also add to 101. The same is true of the third and third-to-last terms. Continuing in this fashion, you soon find yourself with 50 pairs of numbers adding to 101; 50 times 101 is 5,050.

SECOND SOLUTION[1]

A simple technique you can picture easily is the following:

$$\begin{array}{cccccc} 1 & 2 & 3 & \dots & n-1 & n \\ \underline{n} & \underline{n-1} & \underline{n-2} & \dots & \underline{2} & \underline{1} \\ n+1 & n+1 & n+1 & \dots & n+1 & n+1 \end{array}$$

There are n terms each equal to $n+1$. The required sum is half the grand total: $\frac{n(n+1)}{2}$.

THIRD SOLUTION

I read somewhere many years ago that the high school drop-out Albert Einstein devised the following alternative solution technique at age 15. Think of each summand, i, in the sum $\sum_{i=1}^{100} i$ as a group of i marbles in a row from $i=1$ to $i=100$ (see the array following). Stacking each row of marbles on top of each previous row, you get the array including both the diagonal and the lower-triangular off-diagonal. Were the array full, it would contain $100\times 100 = 100^2$ marbles. So, your answer must be roughly half this (roughly 50×100). This is not exact because although the array contains two triangular-shaped off-diagonals (upper and lower), there is only one diagonal. If you add another diagonal, and *then* split the total in two, you get the right answer. The diagonal contains 100 marbles, so the right answer must be $\frac{100^2+100}{2} = 5,000+50$, as before.

[1] I thank Tom Arnold for this solution technique. I am responsible for any errors.

$$\begin{array}{cc} & \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & \cdots & 100 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ \vdots \\ 100 \end{array} & \left(\begin{array}{cccccccc} \bullet & & & & & & & \\ \bullet & \bullet & & & & & & \\ \bullet & \bullet & \bullet & & & & & \\ \bullet & \bullet & \bullet & \bullet & & & & \\ \bullet & \bullet & \bullet & \bullet & \bullet & & & \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \cdots & \bullet \end{array}\right) \end{array}$$

More generally, the sum from 1 to n may be written down as $\frac{n^2+n}{2} = \frac{n(n+1)}{2}$. Just picture the square array of side length n, add another diagonal, and split the total in half.

To calculate $\frac{n(n+1)}{2}$ quickly in your head, note that one of n or $n+1$ must be even and thus divisible by two. You should divide the even number by two and multiply the odd number remaining by the result. In our case,

$$\frac{100\times 101}{2} = \frac{100}{2}\times 101 = 50\times 101 = 5,050.$$

Finally, note that two more solutions appear in the answers to Question 1.41, starting on page 91.

Answer 1.3: This question has been very popular. Sometimes it is golf balls, sometimes marbles, sometimes coins. Most people find it very challenging.[2]

The first step is to split the 12 marbles into three groups of four. Each group of four has two subgroups, a singleton and a triplet: $\{\{1\}_A, \{3\}_A\}$, $\{\{1\}_B, \{3\}_B\}$, and $\{\{1\}_C, \{3\}_C\}$.

Compare $\{\{1\}_A, \{3\}_A\}$ to $\{\{1\}_B, \{3\}_B\}$. If they balance, then the odd ball is in group C. In this case, compare $\{3\}_C$ to $\{3\}_B$. If $\{3\}_C$ is heavier (or lighter), then comparing any two marbles from within $\{3\}_C$ immediately locates the odd one; if $\{3\}_C$ balances $\{3\}_B$, then compare $\{1\}_C$ to $\{1\}_B$ to see whether $\{1\}_C$ is heavier or lighter.

If the initial comparison is unbalanced, say $\{\{1\}_A, \{3\}_A\}$ is heavier than $\{\{1\}_B, \{3\}_B\}$, then rotate groups $\{3\}_A$, $\{3\}_B$, and $\{3\}_C$ and compare grouping $\{\{1\}_A, \{3\}_B\}$ to $\{\{1\}_B, \{3\}_C\}$ (while holding out $\{\{1\}_C, \{3\}_A\}$). If they balance, then a heavy marble is in $\{3\}_A$ and comparing any two marbles from within $\{3\}_A$ immediately locates the odd one. Suppose they do not balance.

[2] I heard about one guy who got home, took 12 golf balls, and tried to solve this by physically manipulating them. I understand that he was still unsuccessful. This particular solution technique combines independent contributions of Juan Tenorio, Bingjian Ni, Yi Shen, and Jinpeng Chang. I am responsible for any errors.

If $\{\{1\}_A, \{3\}_B\}$ is heavy, then either $\{1\}_A$ is heavy, or $\{1\}_B$ is light. Compare $\{1\}_A$ to $\{1\}_C$ to finish. If $\{\{1\}_A, \{3\}_B\}$ is light, then $\{3\}_B$ is light and comparing any two marbles within $\{3\}_B$ immediately locates the light one.

In each case, only three weighings are needed. This technique is generalized in Answer 1.14 (the "90-coin problem").

Answer 1.4: This is cute. You (the bug) cannot fly; you have to walk. You must find the shortest path from corner to corner.

In any world, the shortest path between two points is called a "geodesic." On a spherical world (e.g., the Earth's surface), a geodesic is an arc of a "great circle." A great circle is a circle on the surface of the sphere with diameter equal to the diameter of the sphere. For example, aeroplanes typically follow great circles above the Earth (because it is the shortest path and, therefore, the most fuel-efficient path).

Like a spherical world, the cubic-room world has a two dimensional surface. However, the lack of curvature in the cubic-room world means that the shortest distance between two points must be a straight line rather than an arc of a great circle (in a world without curvature, geodesics are straight lines).

If the cubic room is opened up and flattened out it can be seen that the shortest path is a straight line from one corner to the other. In the un-flattened room, this straight line corresponds to two line segments that meet exactly halfway up one wall-floor or wall-wall boundary. Direct computation using Pythagoras' Theorem[3] reveals that the total path length is $\sqrt{5}$ units.

Answer 1.5: The 10×10 macro-cube question has been very popular. The most common mistake is for people to *count* the number of 1×1 micro-cubes on each face and add them up. Even if you do the mathematics correctly (and most people do not), you miss the whole point.

If you focus on the 1×1 micro-cubes on the faces and how to count them directly (e.g., How many faces? How many on each face? How many edges?), you miss the point. Go back now and figure out a better way. As I stated before, the path of greatest resistance bears the highest rewards, so read no further unless you did it a better way.

You must look for structure in a problem that leads you to a simple and speedy solution. The most structure here is to be found in the macro-cube you start with and the (now slightly smaller) macro-cube that remains. The difference between their volumes is how many micro-cubes fell.

[3]Recall Pythagoras' Theorem. Consider a triangle with side lengths X, Y, and Z. If the angle between the sides of length X and Y is $90°$, then it is a "right-angle" triangle. The side of length Z (the "hypotenuse") must be the longest side, and it must be that $X^2 + Y^2 = Z^2$. In this case, the path is the hypotenuse of a triangle of side lengths 2 and 1 in the flattened-out room or two hypotenuses of triangles each of side lengths 1 and $\frac{1}{2}$ in the un-flattened room. In either case, the path is of total length $\sqrt{5}$.

The volume of a cube of side length n is n cubed; that is, n^3. The answer is, therefore, $10^3 - 8^3$.

How do you calculate this without a calculator? You should know that 10^k is a 1 with k zeroes attached, so $10^3 = 1,000$. You should know that $8 = 2^3$ and, therefore, that $8^3 = 2^{3\times3} = 2^9$. You should definitely know that 2^{10} is 1,024. Thus, 2^9 is half of 2^{10} and, therefore, equal to 512. The answer is $1,000 - 512 = 488$. A common mistake is for people to think the answer is $10^3 - 9^3 = 271$, because only "one layer" fell off (you should of course know what 9^3 is also without having to work it out).

Answer 1.6: This is a good question. People tend to overlook the brilliantly simple situation described. If you did any mathematics whatsoever, you probably missed the point.

No calculation is needed to see that at each stage an equal number of male babies and female babies are expected to be born. The proportions of male and female children are, therefore, expected to remain equal at 50%.

Still stuck? Here are the details (assuming equal numbers of boys and girls are born): by the end of the first year, the 100,000 families have 50,000 boys and 50,000 girls. The proportion of male children stands at 50%. By the end of the second year, half of the 100,000 families (the ones without a son) have another child. This adds 25,000 boys and 25,000 girls. There are now 75,000 boys and 75,000 girls. The proportion of male children still stands at 50%. There are still 25,000 families without a son. They add another 12,500 boys and an equal number of girls, and so on.

Some people are tempted to suppose that because all large families have many daughters and a single son, there must be more girls than boys. However, there are not many large families.[4]

Answer 1.7: I like this one. People have given me answers to this one ranging from $0°$ to $75°$ (and many answers in between). The big hand is on the three; the little hand is one-quarter of the way between the three and the four. The answer must be one-quarter of one-twelfth of $360°$. That is, one-quarter of $30°$. That is, $7.5°$ (or $\frac{\pi}{24}$ radians if you like measuring angles in radians).[5]

You should focus on what you know (the angle is non-zero, the big hand is on the three, one hour is one-twelfth of the full circle, and 15 minutes is one-quarter of one hour) and make sure that your answer accords with intuition. For example, if you get $75°$, then something is wrong with your reasoning (or you have never owned an analogue wristwatch).

[4]In fact, the average number of children per family is only $\sum_{k=1}^{\infty} \frac{k}{2^k} = 2$ (obtained using basic probability theory and the following algebraic result derived by me: $\sum_{k=1}^{\infty} \frac{k}{x^k} = \frac{x}{(x-1)^2}$, for $|x| > 1$).

[5]There are 2π radians in a full circle. Thus, $360° = 2\pi$ radians; $180° = \pi$ radians; $90° = \frac{\pi}{2}$ radians; and so on. It is just another way of measuring angles.

Story: 1. He whistled when the interviewer was talking. 2. Asked who the lovely babe was, pointing to the picture on my desk. When I said it was my wife, he asked if she was home now and wanted my phone number. I called security.

Interview Horror Stories from Recruiters
Reprinted by kind permission of *MBA Style Magazine*
©1996–2009 MBA Style Magazine, `www.mbastyle.com`

Answer 1.8: I present both a brute-force approach (with some algebra) and an elegant approach looking at the bigger picture.

FIRST SOLUTION

By the time the minute hand gets to the three, the hour hand will have moved on a little. There is then a "catch-up time" of a minute or two after 3:15, so the hands must be coincident around 3:16 or 3:17. As time passes from three o'clock to four o'clock, the minute hand whips 60 minutes around the entire face, and the hour hand moves slowly from pointing at the 3 to pointing at the 4 (marked off as five increments of one minute on my watch). At all times, the proportion of the full 60 minutes traversed by the minute hand equals the proportion of the five increments of one minute traversed by the hour hand. Let M minutes after 3PM be the first time at which the hands are coincident, then our argument above implies directly that the proportion of the full 60 minutes covered by the minute hand equals the aforementioned catch-up time as a proportion of the five increments of one minute

$$\begin{aligned} \frac{M}{60} &= \frac{M-15}{5} \\ &= \frac{M}{5} - 3 \\ \Rightarrow \frac{M}{5} - \frac{M}{60} &= 3 \\ \Rightarrow \frac{11M}{60} &= 3 \\ \Rightarrow M &= \frac{180}{11} = 16\frac{4}{11}. \end{aligned}$$

Thus, the hands are coincident at 3:16 and $\frac{4}{11}^{th}$'s of a minute. That is 3:16:21.82.

SECOND SOLUTION

At high noon, the hands are coincident. At midnight, the hands are coincident. The time between noon and midnight is cut into 11 equally-spaced time intervals of one-eleventh of 12 hours (1:05:27.27). At the end of each of these intervals, the hands are coincident. For the question at hand, the answer is three times one-eleventh of 12 hours: 3:16:21.82. The full set of "coincident times" are as follows:

$$\left(\begin{array}{r} 1:05:27.27 \\ 2:10:54.55 \\ 3:16:21.82 \\ 4:21:49.09 \\ 5:27:16.36 \\ 6:32:43.64 \\ 7:38:10.91 \\ 8:43:38.18 \\ 9:49:05.45 \\ 10:54:32.73 \\ 12:00:00.00 \end{array}\right)$$

Answer 1.9: Well now, this looks pretty complicated the first time you see it. However, there is a simple way to figure it out. If you think about it, you see that the only brokers who touch the switch for light bulb number 64 are those whose numbers are divisors of 64.

That is, light bulb number 64 has its state of illumination changed by brokers whose numbers are factors of 64. That is, brokers 1, 2, 4, 8, 16, 32, and 64 flip the switch. Because light bulb 64 is originally *off*, it must be that after this odd number of switches it is *on*. See Answer 1.10 for a closely related but more general solution technique.

Answer 1.10: If you now know the answer to Question 1.9, you should be able to figure this one out swiftly. If you have not yet figured out Question 1.9, then read no further—solve that one first.

The only way for a light bulb to be illuminated after the 100th person has passed through is if its switch was flipped an odd number of times. The switch for light bulb number K gets flipped only by people whose numbers are factors of K. Thus, the only light bulbs illuminated at the conclusion are those with a number that has an odd number of factors.

However, factors for numbers go in pairs. For example, 32 has factors $(1,32)$, $(2,16)$, and $(4,8)$. This means that 32 has an even number of factors, and bulb 32 is not illuminated at the conclusion. In fact, at first glance, all numbers have an even number of factors.

However, you do get an odd number of factors if two factors (one pair) are identical. For example, 64 has $(8,8)$ as one pair. If one pair of factors are identical, then the original number must be a "perfect square." Therefore, the only numbers with an odd number of factors are the perfect squares.

There are exactly 10 perfect squares between 1 and 100, and they are 1, 4, 9, 16, 25, 36, 49, 64, 81, and 100 (i.e., 1^2, 2^2, 3^2, $\ldots$ 10^2). These are the numbers of the 10 bulbs that are illuminated after the 100th person has passed through the room.

Answer 1.11: This is an old favorite. I have tried this out on people and have received almost all imaginable responses. The answer is three, and it cannot be anything else. Two socks can be different, but a third must match one of the first two—giving a matching pair.

Answer 1.12: You get the answer by working backwards. If I am your opponent, and I am able to call out "39," then you cannot reach 50, but I can after you say whatever you say. So, my goal is to call out "39." However, if I am able to call out "28," then you cannot get to 39, but I can after you say whatever you say. So, my goal is to call out "28." To get to 28, I need only to be able to call out "17," and to do this, I need only to be able to call out "6."

So, my strategy, as your opponent, is to get onto the series 6, 17, 28, 39, 50 at whichever point I can. If you get to go first, you should call out "6." As long as you know the winning numbers and stick to them, you cannot lose. If you start with anything other than 6, I cannot lose.

Answer 1.13: Safe-cracking in a finance interview? Yes indeed. The naive answer is that there are 40 possible numbers for the first combination, 40 for the second, and 40 for the third. It would then take at most 40^3 possible trials to get the safe open. That is 64,000 trials. There are two factors that reduce this number considerably. The first you should have figured out; the second you are excused.

The first factor is that although three numbers are required to open the safe, you need only find the first two of them. If you dial the first two numbers correctly, then you need only turn the dial until the safe pops open. You do not need to know the last number. This gives 40^2 possible combinations. That is only 1,600 trials. For extra safe-cracking advice along these lines see Gleick (1993, pp189–190).[6]

The interviewer in this case suggested a second factor, as follows (and I think it is a little unfair to any interviewee who is inexperienced in safe-cracking). The safe is a mechanical device designed with a particular tolerance for inaccuracy. If the first combination number is 14, then dialing either 13, 14, or 15 suffices. This tolerance for inaccuracy brings you down to roughly $\left(\frac{40}{3}\right)^2$ trials. This is fewer than 200 trials.

Answer 1.14: To minimize the maximum possible cost of weighing, your strategy must use the scales as few times as possible, wherever the location of the "bad" coin. From Answer 1.3, you know that you may need as many as three weighings to find a bad coin in a group of 12. You have 90 coins, so it must take at least four weighings. However, by the same argument, if you had 144 coins (12 groups of 12), you could identify a bad group of 12 in three weighings

[6] I went to a presentation at MIT at which Jim Gleick (pronounced "Glick") talked about his then soon-to-be-published book "Genius." He talked about genius in general and Richard P. Feynman in particular. Feynman was an interesting guy, and this is a good book about him.

and then the bad coin in another three. So, (because 90 is less than 144) it should take no more than six weighings—either four, five, or six weighings.

In fact, it takes only five weighings (and at most $500) to both find the bad coin and identify it as heavy or light. I present two quite different solution techniques plus a third quasi-solution: the first is an ingenious "hammer-and-tongs" technique, the second is slightly more structured, and the third generalizes the second but applies only in special cases. In each case, it takes only five weighings to both find the bad coin and identify it as heavy or light.

FIRST SOLUTION[7]

The technique here is similar to the solution technique to Question 1.3—be sure to answer that question before answering this one. The first move is to divide the 90 coins into three groups of 30. Weigh two of the groups of 30 coins. Either the scales balance, or they do not. If the scales balance, then you are left with one group of 30 containing the bad coin. You may "hold out" 10 of these 30 and compare the remaining 20—with 10 on each side. If the scales balance, you get one group of 10 containing the bad coin. If they do not balance, you have one group of 10 coins potentially heavy and one group of 10 coins potentially light. Stop here if you just wanted to know how to start the solution process. This should be enough for you to finish.

Return for a moment to the case in which the two groups of 30 do not balance. Set aside 10 potentially heavy coins and 10 potentially light coins. Take the remaining 20 potentially heavy and 20 potentially light coins and swap 10 of them from one side of the scales with 10 of them from the other side of the scales, keeping track of which were swapped and which stayed put. Whether they balance or not, you can immediately identify one group of 10 coins that is potentially heavy and one group of 10 coins that is potentially light—the other 40 are "good" coins.

Thus, after two weighings, the problem reduces either to one group of 10 coins containing the bad coin (no further information) or two groups of 10 coins (where one group potentially contains a heavy coin, and the other potentially contains a light one). I need only demonstrate the solution technique for each case.

Suppose you have 10 coins, and one of them is bad. You can find the bad one in three weighings simply by adding two good coins and following Answer 1.3 for the 12-ball case. This finds you the bad coin in a total of five weighings.

Suppose instead that you have the two groups of 10 coins (where one group potentially contains a heavy coin, and the other potentially contains a light one). Use the notation "3↑" to denote three potentially light coins, "3↓" to denote three potentially heavy coins, and "good" to denote one coin known to be neither heavy nor light. In this case, you begin at the end of the second

[7] I thank Eva Porro (then at the Universidad Complutense de Madrid) for this solution technique. I am responsible for any errors.

weighing with {10↑} and {10↓} on the scales. Hold out 3↑ and 3↓ coins and place the following on the scales for weighing number three: {3↑,4↓} versus {3↓,4↑}.

Suppose the scales balance with {3↑,4↓} versus {3↓,4↑}. Then you have 3↑ and 3↓ coins left. Hold out 1↑ and 1↓ and weigh {1↑,1↓} versus {1↓,1↑}. If these balance, weigh the hold out 1↑ against one good coin to find the bad one; if they do not balance, you get 1↑ and 1↓ from the light and heavy sides, respectively; and you need only compare one of them to a good coin to find the bad one. This gives a total of five weighings in either case.

Suppose the scales do not balance with {3↑,4↓} versus {3↓,4↑}. If the first group appears lighter, then you get 3↑ and 3↓ coins as in the previous paragraph and able to be solved in a total of five weighings. If the second group appears lighter, then you get 4↓ and 4↑ coins. This is just like the first weighing of two groups of four in the 12-ball problem in Question 1.3, and you know the bad coin can be identified in only two more weighings by rotating "triplets." In each case, the bad coin is both located and identified as heavy or light in only five weighings.

> **Story:** One of my students was told "Take your jacket off—it's going to get hot in here."

SECOND SOLUTION[8]

Begin by noting that if you have a group of 3^k coins that is known to contain a heavy coin, it takes only k weighings to identify it. You can see this as follows: split the group of 3^k coins into three subgroups each of size 3^{k-1}; now compare any two subgroups on the scales. Whether the scales balance or not, you know immediately which of the three subgroups contains the heavy coin. It thus takes only one weighing to go from a group of 3^k coins known to contain a heavy coin to a group of 3^{k-1} coins known to contain a heavy coin. Proceeding in this fashion, it takes k weighings to go from a group of 3^k coins known to contain a heavy coin to a single coin known to be heavy. The same result applies if the initial sample is known to contain a light coin.

Therefore, if you know that the bad coin in your sample is heavy (or if you know that it is light), table A.1 gives the correspondence between sample size and number of weighings required to locate the bad coin. I now use table A.1 to answer the question. Begin by splitting the sample into as few groups of form 3^k as possible.[9] In this case, $90 = 81 + 9$, so you choose one group of 81 and one group of nine. Split the group of 81 into three subgroups of 27. Call these groups $\{27\}_A$, $\{27\}_B$, and $\{27\}_C$. Now use the scales to compare groups $\{27\}_A$ and $\{27\}_B$. Now use the scales again to compare groups $\{27\}_A$ and $\{27\}_C$. If the bad coin is in the group of 81, then these two weighings are

[8] I thank Bingjian Ni for this solution technique. I am responsible for any errors.

[9] Can you make a conjecture about the sample size, its ternary (i.e., base three) representation, and the number of weighings needed to find the bad coin/marble?

Table A.1: Weighings Needed to Find Bad Coin

Sample Size	Weighings (if bad coin is known heavy)
3	1
9	2
27	3
81	4
243	5
⋮	⋮

Note: If you have a sample of coins and you know that there is a bad coin in your sample and that it is heavy (or if you know that it is light), then the table gives the number of weighings required to locate the bad coin.

sufficient to identify which subgroup of 27 the bad coin falls into and whether it is heavy or light. Consulting table A.1, you can see that in this case it takes only three additional weighings to find the bad coin.

If both the initial weighings balance (i.e., $\{27\}_A$ versus $\{27\}_B$ and $\{27\}_A$ versus $\{27\}_C$ both balance), then the bad coin is in the group of nine. Compare the group of nine to nine good coins taken from the group of 81. This tells you whether the bad coin is heavy or light. Consulting table A.1, you can see that in this case, it takes only two more weighings to find the bad coin. Alternatively, you could have split the group of nine into three groups of three and weighed two pairs of them. This identifies the group of three containing the bad coin and tells you whether it is heavy or light. Consulting table A.1, you can see that in this case, it takes only one more weighing to find the bad coin. In each case, the bad coin is both located and identified as heavy or light in only five weighings (at a maximum cost of $500).

THIRD SOLUTION[10]

Suppose you are given $N = \frac{3^n-3}{2}$ balls for some positive integer n. The balls appear identical, but one ball is odd—either heavy or light; you do not know which. Then it takes n weighings to both find the odd ball and identify it as heavy or light (see table A.2).

It is no coincidence that the first column in table A.2 is the partial sums of the first column in table A.1—this technique generalizes the second. I prove the particular case $N = 120$ (i.e., $n = 5$), but the proof generalizes directly to any $N(n) = \frac{3^n-3}{2}$.

Put the 120 balls into three groups of 40. Each group of 40 is a cohort of subgroups of size 3^k for $k = 0$ to $k = n - 2$: $\{\{1\}_A, \{3\}_A, \{9\}_A, \{27\}_A\}$;

[10] I thank Yi Shen for this solution technique. I am responsible for any errors.

Table A.2: Weighings Needed to Find Bad Coin

Balls Supplied $N = \frac{3^n-3}{2}$	Weighings Needed n
3	2
12	3
39	4
120	5
363	6
$\vdots$	$\vdots$

Note: If you have a sample of coins and you know that there is a bad coin in your sample but not whether it is heavy or light, then the table gives the number of weighings required to locate the bad coin.

$\{\{1\}_B, \{3\}_B, \{9\}_B, \{27\}_B\}$; and $\{\{1\}_C, \{3\}_C, \{9\}_C, \{27\}_C\}$.

Compare cohorts A and B. If they balance, then you have 80 good balls, and cohort C contains the bad ball. In this case, compare $\{27\}_C$ to the good balls $\{27\}_A$. If $\{27\}_C$ contains the bad ball, then table A.1 says you need three more weighings. If $\{27\}_C$ is good, then compare $\{9\}_C$ to the good balls $\{9\}_A$. If $\{9\}_C$ contains the bad ball, then table A.1 says you need two more weighings. If $\{9\}_C$ is good, then compare $\{3\}_C$ to the good balls $\{3\}_A$. You need only one more weighing—either because $\{3\}_C$ is bad (see table A.1), or because $\{3\}_C$ is good (thus $\{1\}_C$ is bad).

If the initial comparison of A and B does not balance, then rotate (like changing car tires) groups $\{27\}_A$, $\{27\}_B$, and $\{27\}_C$ and compare grouping $\{\{1\}_A, \{3\}_A, \{9\}_A, \{27\}_B\}$ to $\{\{1\}_B, \{3\}_B, \{9\}_B, \{27\}_C\}$ while holding out $\{\{1\}_C, \{3\}_C, \{9\}_C, \{27\}_A\}$. If they balance, $\{27\}_A$ contains the bad ball, and it is known to be heavy or light, and table A.1 says three more weighings are needed. Otherwise, the scales tilt the same way, you can discard the $\{27\}$'s, rotate the $\{9\}$'s, and compare $\{\{1\}_A, \{3\}_A, \{9\}_B\}$ to $\{\{1\}_B, \{3\}_B, \{9\}_C\}$ while holding out $\{\{1\}_C, \{3\}_C, \{9\}_A\}$. You either find $\{9\}_A$ is bad, known heavy or light, and apply table A.1, or you discard the $\{9\}$'s and rotate the $\{3\}$'s. Continue reducing the problem until convergence—in at most five weighings.

One problem with this method is what to do if given only 90 balls (more than 39 but fewer than 120). I guess you ask for 30 extra ones and then follow the procedure for 120.

Answer 1.15: This question is unusually esoteric, but I like it. The result is known as Liouville's Theorem. It can be proved directly using Picard's The-

orem,[11] or with very slightly more work, using a Cauchy integral.[12] I have chosen, however, to prove it from first principles.[13]

Begin by proving a lemma (a "helping theorem" to be used in a later proof). The lemma is used in the proof of a theorem that answers the interview question. If you have a mathematical background but cannot answer the question, you should read the statement of the lemma, and stop reading there. You should then try to complete the remainder of the proof on your own. This is more satisfying than seeing the full proof.

LEMMA: Maximum Modulus of Coefficients of a Power Series[14]
Suppose that $f(z)$ is analytic in the disc $|z| \leq r < \infty$. Let $M(r) \equiv \max\{|f(z)| : |z| \leq r\}$. Then the coefficients a_p in the power series expansion $f(z) = a_0 + a_1 z + \cdots + a_p z^p + \ldots$ satisfy the following bound:

$$|a_p| \leq \frac{M(r)}{r^p} \quad \text{for } p = 1, 2, 3 \ldots$$

PROOF OF LEMMA
With $f(z) = a_0 + a_1 z + \cdots + a_p z^p + \ldots$ in $|z| \leq r < \infty$, divide by z^p to get

$$\frac{f(z)}{z^p} = a_0 z^{-p} + a_1 z^{-p+1} + \cdots + a_p + \ldots.$$

Now change to polar coordinates. Hold $r = |z|$ constant, and integrate $\frac{f(z)}{z^p}$ from $\phi = 0$ to $\phi = 2\pi$ [where $\phi \equiv arg(z)$]:

$$\int_0^{2\pi} \frac{f(z)}{z^p} d\phi = \int_0^{2\pi} a_0 z^{-p} d\phi + \cdots + \int_0^{2\pi} a_p d\phi + \ldots$$

The integral is convergent because, for fixed r and varying ϕ, the original series converges uniformly. Each term in the series expansion contributes an integral of form

$$a_{k+p} \int_0^{2\pi} z^k d\phi.$$

[11] Picard: A non-constant entire function assumes every complex value, with at most one possible exception. Thus, a bounded function must be a constant.

[12] Cauchy integral: let $C(z_0, r)$ be a circle or radius r about arbitrary $z_0 \in \mathbb{C}$, then $f'(z_0) = \frac{1}{2\pi i}\int_{C(z_0,r)} \frac{f(z)}{(z-z_0)^2} dz$ (a contour integral), so $|f'(z_0)| \leq \frac{1}{r} \sup_{z \in \mathbb{C}} |f(z)|$. Bounded f implies the RHS tends to zero as $r \to \infty$. Thus, $f'(z_0) \equiv 0$, for arbitrary $z_0 \in \mathbb{C}$, and f must be a constant.

[13] I thank Naoki Sato and Thomas C. Watson for suggesting the Picard Theorem and Cauchy integral approaches, respectively. Any errors are mine.

[14] This lemma and its proof are adapted from Holland (1973, pp9–10), with copyright permission from Academic Press.

However, for $k \neq 0$ this integral contributes zero:

$$\begin{aligned} a_{k+p}\int_0^{2\pi} z^k d\phi &= a_{k+p}\int_0^{2\pi} r^k[\cos(k\phi) + i\ \sin(k\phi)]\ d\phi \\ &= a_{k+p}\ r^k\left[\frac{\sin(k\phi)}{k} - i\ \frac{\cos(k\phi)}{k}\right]_0^{2\pi} \\ &= 0 \end{aligned}$$

Thus, the only term that contributes anything to $\int_0^{2\pi}\frac{f(z)}{z^p}d\phi$ is $\int_0^{2\pi} a_p d\phi = 2\pi a_p$, (when $k = 0$). It follows that

$$a_p = \frac{1}{2\pi}\int_0^{2\pi}\frac{f(z)}{z^p}d\phi \quad \text{for} \quad p = 0, 1, 2, \ldots .$$

Now, with $M(r) = \max\{|f(z)| : |z| \leq r\}$, and $|z| = r$, it follows that for integer $p > 0$, you get

$$\begin{aligned} |a_p| &\leq \frac{1}{2\pi}\int_0^{2\pi}\frac{|f(z)|}{|z|^p}\ d\phi \\ &= \frac{1}{2\pi r^p}\int_0^{2\pi}|f(z)|d\phi \\ &\leq \frac{1}{2\pi r^p}\int_0^{2\pi}M(r)d\phi \\ &= \frac{M(r)}{2\pi r^p}\int_0^{2\pi}d\phi \\ &= \frac{M(r)}{r^p} \quad \square \end{aligned}$$

I now present the interview question as a theorem and use the previous lemma in its proof.

THEOREM: Bounded Entire Function[15]
If $f(z)$ is entire and bounded, then $f(z)$ is a constant.

PROOF OF THEOREM
Assuming that $f(z)$ is entire implies that $f(z)$ is analytic in the entire finite complex plane. Thus, the Taylor series $f(z) = \sum_{n=0}^{\infty} a_n z^n$ converges for all $|z| < \infty$. If the stated bound (of the theorem) is $|f(z)| \leq M$, say, then from the lemma it follows that

$$0 \leq |a_n| \leq \frac{M(r)}{r^n} \leq \frac{M}{r^n} \quad \text{for all } n > 0 \text{ and all } r.$$

If you let $r \to \infty$, then $0 \leq |a_n| \leq 0$ for all $n > 0$. Thus, $a_n = 0$ for all $n > 0$, and $f(z) = a_0$, a constant. $\square$

[15]This theorem and its proof are adapted from Holland (1973, p10), with copyright permission from Academic Press.

Answer 1.16: After 100 seconds you can be sure that the ants have all walked off the ruler. The answer is the same as if the ants had perfect vision. The key is that if two ants who collide immediately about face and continue, then each member of any colliding pair effectively exchanges its exit route with the other. It is just as if one of the colliding pair crawled over the other and they both kept going without pause.

Answer 1.17: This is a nice question. If you can figure out the correct relationship between the eight lily pads and the single one, you get the answer. If you do not have it yet, or you think it is 3.75 days, then you should stop reading now, and go back and try again. I am serious; this is a nice question, and you lose a great deal by peeking at the answers to help you out.

The naive answer is that it takes $\frac{30}{8} = 3.75$ days. This is, of course, incorrect. The lily pads in the question all grow at the same rate. This means that you may think of the eight lily pads as being equivalent to one big lily pad. Indeed, when the single lily pad is three days old, it has the same area as the eight lily pads do at time zero. This means that you may think of the eight lily pads as a single lily pad that is three days old. It takes another 27 days for a three-day-old single lily pad to cover the pond, so it also takes 27 days for the eight lily pads to cover the pond.

The interviewee suggested that I use 3,000 days instead of 30 days as the time it takes for the single lily pad to cover the pond. The idea was to make the question more confusing. The problem with this is that, no matter how small the initial lily pad (assuming it is visible to the naked eye), it will cover the surface of the Earth within 100 days and the entire solar system not long after. Within 3,000 days, the universe will be blotted out—such is the power of compound growth.

Answer 1.18: You use the same idea as in the previous lily pad question. Each pad needs to cover $\frac{6{,}000}{27}$ square feet to choke the pond. The size of each pad is 2^N after N days, so you need to solve: $\frac{6{,}000}{27} = 2^N$. The solution is $N = \frac{\log(\frac{2{,}000}{9})}{\log(2)} \approx 7.8$ days.[16]

Without a calculator, you can still do it in your head. You calculate $\frac{6{,}000}{27}$ as approximately 200. You know that 2^8 is 256, which exceeds 200; whereas 2^7 is 128, which falls short, so eight days should do the trick.

Answer 1.19: Decimal pricing was introduced on the New York Stock Exchange in 2001. I have left this question in because there are many people who lived with eighths and sixteenths for most of their working life, and they may be tempted to ask you about it.

Most people stumble a little. Do not memorize all the possible sixteenths before your interview—you have worse things to worry about (much worse).

[16]An interesting aside here is that it does not matter which logarithm function you use. The result is the same regardless of the base.

Add or subtract one-sixteenth to get the requested fraction into quarters or eighths and then compensate for your adjustment.

You should remember that $\frac{1}{8}$ is 0.1250 and deduce from that that $\frac{1}{16}$ is 0.0625 (you should be able to give *any* eighth in decimal form). The fraction $\frac{13}{16}$ is only one-sixteenth away from $\frac{12}{16}$ which is exactly three quarters (0.7500). You need only add 0.0625 to 0.7500 to get 0.8125. Similarly, $\frac{9}{16}$ is only one-sixteenth over one half, and is, therefore, 0.5625.

Answer 1.20: A common question. The naive answer is that the snail climbs a net of two feet per day, so it reaches the 10-foot mark at the end of the fifth day. However, on the morning of the fifth day, the snail starts out at the eight-foot mark (having slid down from the nine-foot mark overnight). Two-thirds of the way through the fifth day, the snail reaches the 10-foot mark and stops because there is no pole left to climb.

Answer 1.21: Here are two answers.[17]

1. Turn Switch #1 on. Wait a while. Then turn it off while simultaneously turning Switch #2 on. Go into the room. The illuminated light corresponds to Switch #2. The warm non-illuminated bulb corresponds to Switch #1. The cold non-illuminated bulb corresponds to Switch #3.
2. Guess. You have a one-in-six chance if they are random. However, light switches are not usually random. If you assume the switches are physically located in an order that relates to the physical placement of the bulbs (as they usually are), then you have a fifty-fifty chance!

Answer 1.22: The only way the first man can know the color of his own hat is if he sees the other two wearing red hats—of which there are only two. However, the first man does *not* know his hat color, so the other two must be wearing either both blue or one red and one blue. The second man, upon hearing the first, knows then that he and the third man are either both wearing blue hats, or one wears a red hat, and one a blue. If he still does not know what color hat he is wearing, it must be because the third man is wearing a blue hat. Why? Well, if the third man wears red, then that pinpoints his own hat as blue since this is the only option left from the choice of either both blue or one red and one blue. Since the second man does not know his hat color, then the third man must be wearing blue. The third man, upon hearing the first two, deduces that his own hat is blue via the same reasoning.

> **Story:** The interviewer got up part way through the interview and walked to the door and opened it. The candidate said, "Are we out of time?" The interviewer replied, "No. Out of interest!"

[17] I thank Dahn Tamir for assistance on this question. I am responsible for any errors.

Answer 1.23: This is a very nice question indeed. You may be looking to the solutions for a hint. My first hint is that, if you are using linear algebra (i.e., solving systems of equations by substitution) then stop that right now. There are nine equations in 10 unknowns, so this will get you nowhere. In fact, there are infinitely many integers that solve the problem statement; we are searching for the smallest such number. My second hint is that you might like to try drawing a picture.

FIRST SOLUTION

My first solution technique begins with the simultaneous equation approach and quickly abandons it. Let X denote the solution. Then I know there exist positive integers X_2, X_3, X_4, ... X_{10}, such that

$$\begin{aligned} X &= 2 \times X_2 + 1, \\ X &= 3 \times X_3 + 2, \\ X &= 4 \times X_4 + 3, \\ X &= 5 \times X_5 + 4, \\ &\vdots \\ X &= 10 \times X_{10} + 9. \end{aligned}$$

Looking at the equations, it is clear that the coefficients on the right-hand side differ from the remainders by only one. If we simply add one to both sides of each equation, then the coefficients and the remainders will be identical, and we can collect terms to obtain the following:

$$\begin{aligned} X+1 &= 2 \times X_2', \\ X+1 &= 3 \times X_3', \\ X+1 &= 4 \times X_4', \\ X+1 &= 5 \times X_5', \\ &\vdots \\ X+1 &= 10 \times X_{10}' \end{aligned}$$

Where $X_n' \equiv X_n+1$ for each n between 2 and 10. With this simple restatement, the problem now requires that we find the smallest number X, such that $X+1$ is perfectly divisible by 2, 3, 4, 5, 6, 7, 8, 9, and 10. That is, find a number X, such that $X+1 = LCM(2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10)$, where $LCM(\cdot)$ is the lowest common multiple operator. Given various redundancies, we conclude that $X = LCM(6,\ 7,\ 8,\ 9,\ 10) - 1 = 2520 - 1 = 2519$ is the solution.

Looking at my restatement of the problem, it should be clear that if X solves $X + 1 = K \times LCM(6,\ 7,\ 8,\ 9,\ 10)$, for any positive integer K, then X is also a solution to the problem (but not the smallest unless $K = 1$). That is, $X = K \times LCM(6,\ 7,\ 8,\ 9,\ 10) - 1 = K \times 2520 - 1$ is also a solution for any positive integer K.

I think the interviewer would have been perfectly happy to hear that $X = LCM(6, 7, 8, 9, 10) - 1$, without your having to find the LCM. However, this does leave one question unanswered: What is the most efficient way to find the lowest common multiple of a group of numbers?[18]

SECOND SOLUTION

I did not discover the first solution by looking at the simultaneous equations and using algebra; I discovered it by drawing a picture. It is somewhat difficult to reproduce my picture, but here is an attempt using a sporting analogy (see figure A.1).

I am searching for a number X with the divisor/remainder properties described. I have nine runners to help me: Mr. 2, Mr. 3, ... Mr. 9, and Mr. 10. They are assembled at the start of an arbitrarily long, dead-straight, sand-covered, nine-lane racetrack that has distances measured in meters, beginning at "0" at the starting line (bear with me on this). Like many race tracks made for people, the people do not all start in the same place; their positions are staggered (which makes no sense for a straight track in the real world). Mr. 2 starts 1 meter from the zero line. Mr. 3 starts 2 meters from the zero line. Mr. 4 starts 3 meters from the zero line, ... and Mr. 10 starts 9 meters from the zero line.

The gun fires, and the race begins. Each Mr. n runs taking steps of length n meters (for n between 2 and 10). The runners each leave footprints in the sand on the track. Let them run for a very long time and then look at the footprints (we do not care who wins). Starting at the zero line, the divisor/remainder properties of X imply that the first time you find a row of nine footprints adjacent to each other must be after X meters. The first time the footprints (the solid bullets in figure A.1) are aligned vertically, is when they have reached the solution, X meters from the zero starting line. We can see that the number of meters they step out before beginning is one less than their step size when they run. If you look one pace backwards from the start line (back to position -1 on the race track), then it should be clear that the distance from the -1 position to X (i.e., $X + 1$) is just the LCM of the step sizes taken by the runners (how else could the footprints be adjacent?). It should also be clear that if you look beyond X, you will find another set of adjacent footprints after you travel another LCM meters. This solution is identical to the first.

Answer 1.24: This is easier than it sounds. You do not need any infinite sums, and, if you used them, go back and try again before reading on. For every two miles covered by the first motorcyclist, the second covers three miles. Two plus three is five, and there are five multiples of five between them. This means they will meet after the first has traveled 10 miles and the second 15.

[18]In this case, if you factor each of 6, 7, 8, 9, and 10, you get 2×3, 7, 2×4, 3×3, and 2×5. The LCM, when factorized, must include each of these expressions. Indeed, $2 \times 3 \times 3 \times 4 \times 5 \times 7 = 2{,}520$—the LCM.

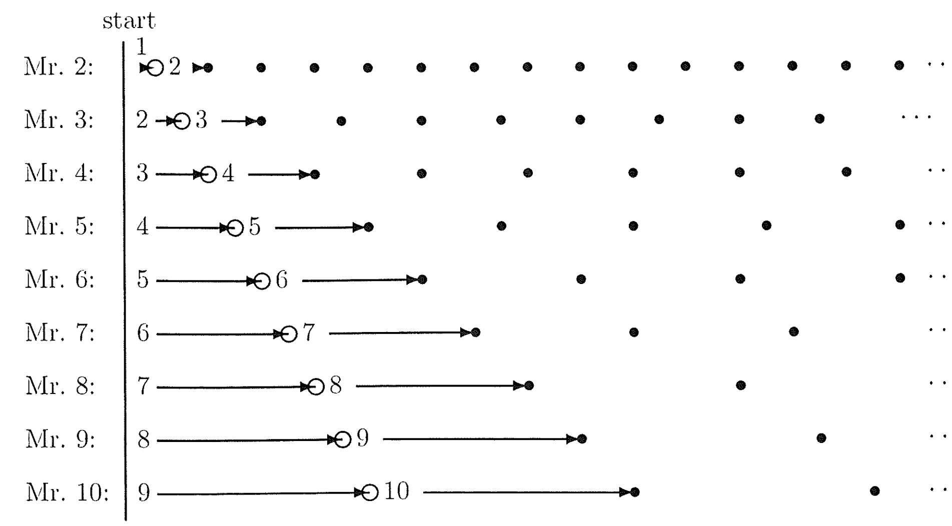


Figure A.1: A Road Race Analogy for the LCM Problem

Note: In the picture, the nine people run side by side. Mr. 2 steps out one meter to start (the hollow bullet) and then take steps of length two meters (his footsteps are solid bullets). Mr. 3 steps out two meters to start and then take steps of length three, and so on up until Mr. 10, who steps out nine meters to start and then takes steps of length 10.

We know that the fly moves at twice the speed of the first motorcyclist, so it must cover 20 miles before its miserable life ends.

Answer 1.25: Let me begin by repeating the constraints:[19]

$$\begin{aligned} A+B+C+D &= 20, \\ B+C+D+E+F &= 20, \\ D+E+F+G+H &= 20, \text{ and} \\ F+G+H+I &= 20 \end{aligned}$$

We have four equations in nine unknowns. The additional information (A to I are some permutation of the integers 1 to 9) restricts the solution space, but there can be no unique solution.

If the first four (A to D) and the last four (F to I), each add to 20, then because $\sum_{i=1}^{i=9} i = 45$, it follows immediately that $E = 5$. If we subtract the second constraint from the first and use $E = 5$, we get $A = F + 5$. If we subtract the fourth constraint from the third, we get $I = D + 5$.

[19]I thank Dahn Tamir and James Hirschorn for contributions to this solution technique. I am responsible for any errors.

The derived restrictions $A = F + 5$, and $I = D + 5$ imply that F and D must be in the set $\{1, 2, 3, 4\}$. Once F and D are chosen, A and I are determined within the set $\{6, 7, 8, 9\}$. There are thus $4 \times 3 = 12$ possible permutations for F, D, A, and I (that is four choices for F followed by three choices for D; see example below). This leaves B, C, G, and H floating in the remaining four spaces. However, subtracting the second constraint from the third implies that $B + C = G + H$. There are four choices for B, but, once B is chosen, C is uniquely determined; see example below. There are thus $12 \times 4 \times 2 = 96$ different solutions.

For example, if $F = 3$ and $D = 4$, then $A = 8$ and $I = 9$ immediately. That leaves B, C, G, and H floating in the remaining four spaces: $\{1, 2, 6, 7\}$. If $B = 1$, then C must equal 7; there is no other choice for C that satisfies $B+C = G+H$. With B and C chosen, there are two ways to allocate G and H to the remaining two slots. In this example, it would either be $G = 2, H = 6$ or $G = 6, H = 2$.

Here are several solutions (reverse the orderings to get several more):[20]

6 8 4 2 5 1 3 9 7,
6 8 4 2 5 1 9 3 7,
6 4 8 2 5 1 3 9 7,
6 4 8 2 5 1 9 3 7

Answer 1.26: Let A denote the area of a triangle of sides a, b, and c.[21] We may make several statements that apply to any triangle and which are clearly visible in figure A.2:

1. The area A is given by $A = \frac{1}{2}bh$.
2. Pythagoras' Theorem implies that $a^2 = h^2 + d^2$ and $c^2 = h^2 + (b + d)^2$.
3. The first Pythagorean result implies $h^2 = a^2 - d^2$. When the two Pythagorean results are subtracted from each other, they imply $d = \frac{c^2-a^2-b^2}{2b}$.

If we combine the above results, we get:

$$\begin{aligned} A^2 &= \frac{1}{4}b^2h^2 \\ &= \frac{1}{4}b^2(a^2 - d^2) \\ &= \frac{1}{4}b^2\left[a^2 - \frac{(c^2-a^2-b^2)^2}{4b^2}\right] \\ &= \frac{1}{4}b^2a^2 - \frac{1}{16}(c^2-a^2-b^2)^2 \end{aligned}$$

[20]Note: The MATLAB commands `perms` and `unqiue` were useful in checking this answer.

[21]I thank Thomas C. Watson for comments on an earlier version of this proof. Any errors are mine.

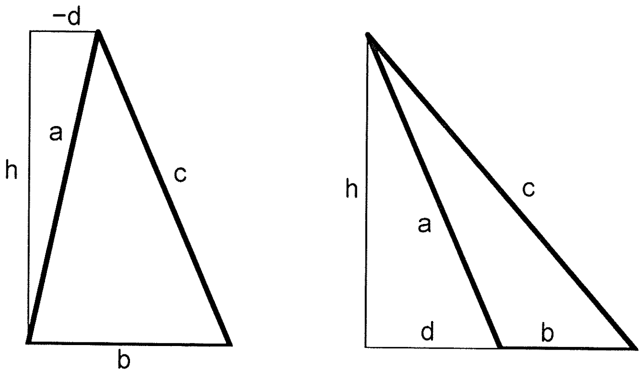


Figure A.2: Two Possible Triangle Configurations

Note: For both triangles configurations, the sides are a, b, and c, and the height is h. The variable d is defined so that $b+d$ measures the distance from the lower right corner to the point where a vertical dropped from the peak touches the base. It follows that $d < 0$ in the first case, and $d > 0$ in the second.

Now, s equals half the perimeter, so $s = \frac{a+b+c}{2}$. It follows that $a = (2s-b-c)$, $b = (2s - a - c)$, and $c = (2s - a - b)$. If we plug each of these into A^2, above, and perform considerable tedious algebra, we arrive at the polynomial[22]

$$A^2(s) = 3s^4 - 4(a+b+c)s^2 + [a^2+b^2+c^2+5(ab+ac+bc)]s^2 + (a+b+c)abc.$$

This may be factored into

$$A^2(s) = [3s - (a + b + c)](s - a)(s - b)(s - c).$$

We have $3s = s+(a+b+c)$, by definition of s, so, $A^2(s) = s(s-a)(s-b)(s-c)$, and thus $A(s) = \sqrt{s(s-a)(s-b)(s-c)}$.

[22]Every term in a polynomial involves positive integral powers of literal numbers (i.e., the letters that represent numbers) pre-multiplied by a factor that does not contain the literals. So, $2x^2y^2 + 5z^3 + 2$ is a polynomial, but $4\sqrt{y}+2$ is not. The "degree" of a polynomial is the degree of the term having the highest degree and non-zero coefficient. The polynomial $2x^2y^2+5z^3+2$ is of degree four $(2+2=4)$. The polynomial $4x^2+2x+1$ is of degree two. The Fundamental Theorem of Algebra says that every polynomial equation of form $f(x) = 0$ (i.e., only one literal) has at least one root (or "zero"). If the polynomial is of degree n, then it has n roots (or zeroes), where repeated roots are counted as often as their multiplicity. The Unique Factorization Theorem says that a factorization of the polynomial $f(x)$ into products of terms of form $(x - \text{root}_i)$ is unique up to trivial sign changes and ordering of terms. See Spiegel (1956) and Spiegel (1981) for more details—both part of the excellent Schaum Outline Series of books.

Story: He took off his right shoe and sock, removed a medicated foot powder and dusted it on the foot and in the shoe. While he was putting back the shoe and sock, he mentioned that he had to use the powder four times a day, and this was the time.

Interview Horror Stories from Recruiters
Reprinted by kind permission of *MBA Style Magazine*

Answer 1.27: I present two solution techniques: a "hammer-and-tongs" brute-force approach and an elegant alternative. I think you should start with a rough guess; mine is about a half.

FIRST SOLUTION[23]

Consider first the simple cases in which there are two or three guests. It soon becomes clear that you need to consider many different overlapping events and that you need to account for intersections of events. That is, you need basic set theory.

Let A_k, for $k = 1, \ldots, N$, denote the event that the kth guest sleeps in the room to which he or she was originally assigned (i.e., his or her "own room"). What we need to find is the probability that *at least one* of the guests ends up in his or her own room. This event is the union of the individual events and occurs with probability: $P(\bigcup_{k=1}^{N} A_k)$.

If you draw the familiar case of three intersecting circles—each representing an event—it is relatively straightforward to deduce the following inclusion-exclusion formula:

$$\begin{aligned} P\left(\bigcup_{k=1}^{N} A_k\right) = & \sum_i P(A_i) - \sum_{1\le i<j\le N} P(A_i \cap A_j) \\ & + \sum_{1\le i<j<k\le N} P(A_i\cap A_j \cap A_k) - \cdots \\ & + (-1)^{N+1} P(A_1 \cap \cdots \cap A_N) \end{aligned}$$

All you are doing here is adding the original event probabilities, then taking out the intersections where you double counted, then adjusting for the fact that you over-compensated, and so on—all of which is very easily seen if you draw intersecting sets for the case $N = 3$. To figure this out, we need to find the probability of the intersection of any group of events. Given the symmetry here, we can, without loss of generality, look at the events in the following order: $1, 2, \ldots, N$.

Given the random allocation of keys, each guest is equally likely to end up in his or her own room. That is, $P(A_i) = \frac{1}{N}$ for any $i \in \{1, 2, \ldots, N\}$. If guest

[23] I thank Taras Klymchuk for suggesting a very similar solution technique. I am responsible for any errors.

1 is given his own key, then guest 2 has a chance of only $P(A_2|A_1) = \frac{1}{N-1}$ of getting her own key back. So, $P(A_1 \cap A_2) = P(A_2|A_1)P(A_1) = \frac{1}{N(N-1)}$. In fact, this result is more general:

$$\begin{aligned} P\left(\bigcap_{k=1}^{m} A_i\right) &= \frac{1}{N(N-1)\cdots(N-m+1)} \\ &= \frac{(N-m)!}{N!} \\ &= \frac{1}{m!\binom{N}{m}}, \quad \text{for any } m \in 1,\ 2,\ \ldots,\ N. \end{aligned}$$

We may now plug this formula for probabilities of intersecting events back into the original inclusion-exclusion formula:

$$\begin{aligned} P\left(\bigcup_{k=1}^{N} A_k\right) &= \sum_{m=1}^{N} (-1)^{m+1} \binom{N}{m} P\left(\bigcap_{k=1}^{m} A_i\right) \\ &= \sum_{m=1}^{N} \frac{(-1)^{m+1}}{m!} \\ &= -1 \times \sum_{m=1}^{N} \frac{(-1)^m}{m!} \\ &= 1 - \sum_{m=0}^{N} \frac{(-1)^m}{m!} \\ &\to 1 - e^{-1} = \frac{e-1}{e} \quad \text{as } N \to \infty. \end{aligned}$$

The final result is about 63%, but a guess of $\frac{2}{3}$ is close enough.

SECOND SOLUTION[24]

Simplify the problem by assuming that the "very large number" of people is almost an infinite number. In this case, it is as though each person is first in line to be allocated a key because the previous finite number of people are negligible compared to the almost infinite number of people waiting to receive keys. It follows that each person has the same probability, $\frac{1}{N}$, of being allocated his or her key. Let X be the number of people who end up sleeping

[24] I thank Jason Roth for supplying this technique. I am responsible for any errors.

in their own rooms, then

$$\begin{aligned} P(X \geq 1) = 1 - P(X = 0) &\stackrel{*}{=} 1 - \left(1 - \frac{1}{N}\right)^N \\ &= 1 - \left(1 + \frac{-1}{N}\right)^N \\ &\rightarrow 1 - e^{-1} = \frac{e-1}{e} \quad \text{as } N \rightarrow \infty, \end{aligned}$$

where "$*$" is true for N infinitely large.

Answer 1.28: First of all, "very small" is classic physics slang for very, very small (i.e., so small that it is a pinpoint mass). If the rock is tossed overboard, the water level falls as though water equal in mass to the mass of the rock is being sucked out of the pool. The rock forces the boat to displace the rock's mass of water. After the rock is gone, the boat rises up, and the water level falls down (Archimedes' Principle).[25]

The next time you are washing dishes, try this experiment. With the sink half-full of water, float a drinking glass. Now drop a steel ball bearing gently into the glass. The glass sinks down, displacing a mass of water equal in mass to the mass of the ball bearing, and the water level rises. Now pluck the ball bearing from the glass, using a magnet. The reverse happens, the glass rises, and the water level falls as though water equal in mass to the mass of the ball bearing is being sucked out of the sink.

Answer 1.29: The answer involves both mathematical induction and game theoretic arguments.

If there is *exactly one* cheating man in the town, Mr. C, say, then every wife except Mrs. C knows who he is. Not only that, but Mrs. C is unaware of any cheats—the stranger's announcement comes as a shock to her. Immediately after the stranger's announcement, Mrs. C asks: "Who can be cheating if I have seen no cheats?" The only possible answer is it is Mr. C. Come the next morning, his happy days are over, and out he goes.

Suppose instead that there are *exactly two* cheating men in town: Mr. C1 and Mr. C2. In this case, Mrs. C1 knows there is one cheat in town (Mr. C2), and Mrs. C2 knows there is one cheat in town (Mr. C1)—the stranger's announcement comes as no shock to either woman. Each thinks there is only one cheat in town and fully expects him to be kicked out the next morning (each wife thinks the other poor woman is in the position of Mrs. C mentioned above). The first morning after the announcement comes, and the streets are bare. Mrs. C1 concludes that Mrs. C2 did not kick her husband out because she did not think he was a cheat. How could Mrs. C2 be so foolish? Mrs. C1 knows

[25] Archimedes said simply that an object in a fluid experiences an upwards force equal to the weight of the fluid that is displaced by the object.

that Mrs. C2 believes the prophecy, so the only possible reason for Mrs. C2 not to have reacted is if Mrs. C2 saw a cheat herself. Mrs. C1 asks herself: "Who did Mrs. C2 see cheating, when the only cheat I can see is Mr. C2?" The only possible answer is that it is her own man, Mr. C1. Come the second morning after the announcement, Mr. C1 and Mr. C2 are both kicked out (the latter because Mrs. C2 went through the same thought process).

Suppose now that there are *exactly three* cheating men in town: Mr. C1, Mr. C2, and Mr. C3. In this case, each of Mrs. C1, Mrs. C2, and Mrs. C3 thinks that there are two cheats in town and believes in the innocence of her own man. However, come the second morning, they are each very surprised to find the streets empty. Had there been exactly two cheats, as each of the wives had surmised, then the cheats should have been kicked into the street two mornings after the announcement—as per the argument above. The empty street means that a third cheater exists—one previously assumed innocent! So, three mornings after the announcement, all three cheating men are bounced out into the street.

More generally, let me assert that if there are exactly n cheats, then they will all be kicked out into the street on the nth morning after the stranger's announcement. If my assertion is true for n cheats, and a wife sees n other cheats but finds the streets bare on the nth morning, then she is shocked to conclude that her own man must be unfaithful to her. She (and each of the other n wives) will kick her man out the next morning. That is, if there are $n+1$ cheats, then they will be kicked out on the $(n+1)$st morning. That is, if my assertion is true for n cheats, then it is also true for $n+1$ cheats.

I proved my assertion to be true for n equal to each of one, two, and three. It follows my mathematical induction that it is true for all n (in fact, I needed only to prove it for $n = 1$ for the proof to go through).

It follows that if cheating men are kicked into the street for the first time on the tenth morning after the stranger's announcement, then there must be exactly 10 of them.

Answer 1.30: This is one of the easier questions in the book. If you are peeking here for a solution, then go back and think about mathematical induction.

Let $V(n)$ denote the minimum number of moves needed for n rings. I *assert* that $V(n) = 2^n - 1$, for all positive integers n (I will justify this shortly). The proof uses mathematical induction.

<u>Case $n = 1$:</u> With one ring, it certainly takes exactly one move. My assertion is thus true for the case $n = 1$.

<u>Case $n = N$:</u> Suppose that my assertion is true for $n = N$, and consider the case $n = N+1$. By assumption, it takes $V(N) = 2^N - 1$ moves to get the first N rings to pole #2. Use one additional move to get ring #$(N+1)$ to pole #3. Now use $V(N) = 2^N - 1$ moves to move the N rings on pole #2 to pole #3. The total number of moves used is $2V(N) + 1 = 2(2^N - 1) + 1 = 2^{(N+1)} - 1$.

However, this is just $V(N+1)$. That is, if my assertion is true for $n = N$, it is also true for $n = (N+1)$.

The result follows, because I showed my assertion is true for $n = 1$, and I showed that if my assertion is true for $n = N$, then it is also true for $n = N+1$. In particular, because I showed the assertion to be true for $n = 1$, it follows that it must be true for $n = 2$. It then follows that because the assertion is true for $n = 2$, it is also true for $n = 3$, ... and so on, up to ∞.[26]

In the particular case of $n = 64$ rings, it takes $V(64) = 2^{64} - 1$ moves. This number is large:

$$V(64) = 2^{64} - 1 = 18,446,744,073,709,551,615.$$

At one move per second, it would take you 584.5 billion (i.e., 584.5 thousand million) years to complete this task. The Earth will fall into the Sun in less than one-hundredth of this time period.

Answer 1.31: The ordinary differential equation (ODE) $u'' + u' + u = 1$ has a simple solution. This is a second-order linear ODE with constant coefficients, so we need only search for solutions to the homogeneous form $u'' + u' + u = 0$, and then tag on a solution to the specific nonhomogeneous equation given.

Solutions to a second-order linear homogeneous ODE of form $Au'' + Bu' + Cu = 0$ are of form[27]

$$u(x) = ae^{\lambda_1 x} + be^{\lambda_2 x},$$

where λ_1 and λ_2 are the roots of the characteristic equation:

$$A\lambda^2 + B\lambda + C = 0.$$

It follows (using the quadratic formula) that

$$\lambda_1, \lambda_2 = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} = \frac{-1 \pm \sqrt{1-4}}{2} = \frac{-1}{2} \pm \frac{\sqrt{3}}{2}i,$$

where $i \equiv \sqrt{-1}$. In our case, $u = 1$ is a solution to the specific nonhomogeneous ODE, so the general solution must be of form

$$u(x) = ae^{\left(\frac{-1}{2} + \frac{\sqrt{3}}{2}i\right)x} + be^{\left(\frac{-1}{2} - \frac{\sqrt{3}}{2}i\right)x} + 1,$$

for arbitrary constants a and b. To pinpoint a and b, you need two initial conditions (not supplied here) in addition to the ODE.

If you prefer a solution with real-valued functions (but with possible complex coefficients), then you may use $e^{(\lambda + \mu i)x} = e^{\lambda x}e^{\mu i x} = e^{\lambda x}[\cos(\mu x) + i\sin(\mu x)]$,

[26] A natural question to ask is how I guessed that $V(n) = 2^n - 1$ to begin with. I got this because I figured that $V(n+1) = 2V(n) + 1$ had to hold, and $V(1) = 1$ is obvious. These together are sufficient to deduce the functional form of $V(n)$.

[27] Unless $\lambda_1 = \lambda_2 = \lambda$, say (i.e., a repeated root), in which case solutions are of form $u(x) = axe^{\lambda x} + be^{\lambda x}$.

and then add and subtract the two components of the general solution to get an equivalent solution (Boyce and DiPrima [1997, p. 148]):

$$u(x) = a'e^{-\frac{1}{2}x}\cos\left(\frac{\sqrt{3}}{2}x\right) + b'e^{-\frac{1}{2}x}\sin\left(\frac{\sqrt{3}}{2}x\right) + 1$$

Answer 1.32: The obvious application is to proportions of a portfolio invested in risky assets (see Questions 3.19 and 3.20). Make the substitution $b = 1 - a$. Then the variance of the sum is

$$V(S) = a^2\sigma_X^2 + 2a(1-a)\rho\sigma_X\sigma_Y + (1-a)^2\sigma_Y^2.$$

The first-order condition is $\frac{\partial V(S)}{\partial a} = 0$. The partial derivative is:

$$\begin{aligned}\frac{\partial V(S)}{\partial a} &= 2a\sigma_X^2 + 2\rho\sigma_X\sigma_Y - 4a\rho\sigma_X\sigma_Y + 2(1-a)(-1)\sigma_Y^2 \\ &= 2\left[a(\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2) + \rho\sigma_X\sigma_Y - \sigma_Y^2\right].\end{aligned}$$

Thus, the particular a that satisfies the first-order condition is

$$a^* = \frac{\sigma_Y^2 - \rho\sigma_X\sigma_Y}{(\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2)}.$$

We should check the second-order condition

$$\left.\frac{\partial^2 V(S)}{\partial a^2}\right|_{a=a^*} > 0,$$

to make sure this is a minimum, not a maximum. This is straightforward here because

$$\begin{aligned}\frac{1}{2}\frac{\partial^2 V(S)}{\partial a^2} &= \sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2 \\ &\geq \sigma_X^2 - 2(+1)\sigma_X\sigma_Y + \sigma_Y^2 \\ &= (\sigma_X - \sigma_Y)^2 \\ &\geq 0,\end{aligned}$$

and the first inequality is strict unless $\rho = +1$.

In fact, I have solved the unconstrained problem—ignoring the constraint $0 \leq a \leq 1$. If a^* breaches the constraints, the constrained solution for a is either 1 or 0, depending upon whether σ_X or σ_Y is the smaller respectively.[28]

[28] The a^* will breach the constraints if the correlation ρ is large enough or the standard deviations are disparate enough that either $\frac{\sigma_X}{\sigma_Y} < \rho$ or $\frac{\sigma_Y}{\sigma_X} < \rho$.

In the special case where $\rho = -1$ (perfect negative correlation), the solution for a^* is given by

$$\begin{aligned} a^* &= \frac{\sigma_Y^2 - \rho\sigma_X\sigma_Y}{(\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2)} \\ &= \frac{\sigma_Y^2 + \sigma_X\sigma_Y}{(\sigma_X^2 + 2\sigma_X\sigma_Y + \sigma_Y^2)} \\ &= \frac{\sigma_Y(\sigma_X + \sigma_Y)}{(\sigma_X + \sigma_Y)^2} \\ &= \frac{\sigma_Y}{(\sigma_X + \sigma_Y)}, \end{aligned}$$

and this particular a^* gives variance of $aX + bY$ equal to zero.

> **Story:** I recall reading in the WSJ about one young woman who was asked her greatest weakness (a common interview question). Without thinking, she blurted out the answer "chocolate!"

Answer 1.33: The lighthouse question is an old favorite.[29] The lighthouse is a distance L from the coast. The beam of light casts a "spot" a distance R across the sea from the lighthouse (see figure A.3).

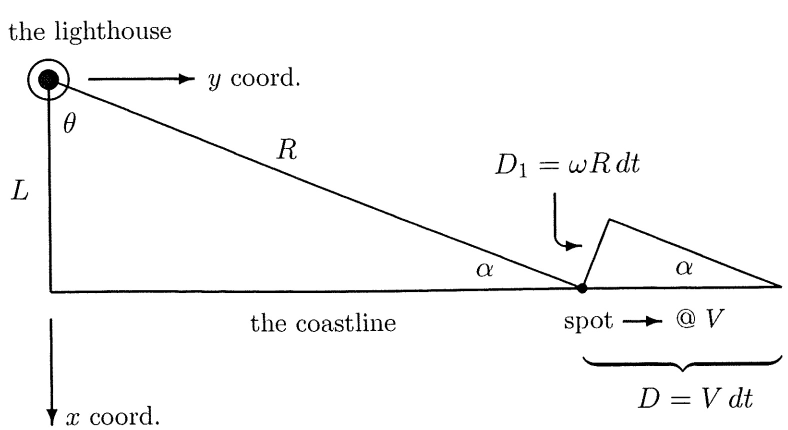


Figure A.3: The Lighthouse Problem

Note: Refer to this figure for both solutions to the lighthouse problem. The first solution uses the x-y coordinates and θ; the second solution uses α, D, and D_1; both solutions use L, R, and V.

[29] I thank Valeri Smelyansky for advice. Any errors are mine.

FIRST SOLUTION

Using the coordinates in figure A.3, the coastline is the line $x = L$. The spot hits the coastline at the coordinate pair $(x, y) = (L, L\tan(\theta))$, where θ is the angle between the beam and the x-axis. Suppose $\theta = 0$ when $t = 0$, then $\theta = \omega t = \frac{2\pi}{60}t$ where $\omega = \frac{2\pi}{60}$ is the angular velocity in radians per second (the beam makes one revolution per minute and t is measured in seconds). The speed V of the spot along the coastline is the partial derivative of $y = L\tan(\theta) = L\tan(\omega t)$ with respect to t:

$$V = \frac{\partial}{\partial t}[L\tan(\theta)] = \frac{\partial}{\partial t}[L\tan(\omega t)] = \omega L\sec^2(\omega t) = \frac{\omega L}{\cos^2(\omega t)}$$

From figure A.3 we see that $\cos(\omega t) = \frac{L}{R}$, so we conclude that $V = \frac{\omega L}{(L/R)^2} = \frac{\omega R^2}{L}$. In our particular case, with $\omega = \frac{2\pi}{60}$ radians per second and $L = 3$, the speed of the beam is $\frac{\pi R^2}{90}$ miles per second. When the beam is $3L = 9$ miles down the coastline, $R^2 = 10L^2 = 90$, and the speed is just π miles per second.

More than once, people have suggested to me that the velocity V is a constant (i.e., V is the same regardless of how far along the coast the spot is cast)—this is clearly incorrect.

SECOND SOLUTION

Follow the beam's course for a small time interval dt. In figure A.3, we see that the beam's spot covers a distance $D = V dt$ along the coast, while the beam's "perpendicular motion" covers a distance $D_1 = \omega R dt$ (where V is the spot's speed along the coast, and $\omega = \frac{2\pi}{60}$ radians per second is the beam's angular velocity). For small dt, the distance triangle is a right-angle triangle, so $\sin(\alpha) = \frac{D_1}{D} = \frac{\omega R dt}{V dt} = \frac{\omega R}{V}$. Looking at the larger triangle, we see $\sin(\alpha) = \frac{L}{R}$. Thus, $V = \frac{\omega R^2}{L}$, as before.

Answer 1.34: There are many different ways to solve this problem. Let me begin with a "hammer-and-tongs" approach using algebra. When you see how neat the solution is, try to come up with an argument that uses no mathematics whatsoever (I present such an argument after the hammer-and-tongs approach). Please see figure A.4.

FIRST SOLUTION

Identify the squares using horizontal and vertical indices, counting from the northwest corner. Let i count down and j count across. Then it is readily seen that the square with coordinates (i, j) has $(i + j - 1)$ cubes on it. It follows that the total number of cubes is given by (in the general case of an $n \times n$

1	2	3	4	⋯	19	20
2	3	4	5	⋯	20	21
3	4	5	6	⋯	21	22
4	5	6	7	⋯	22	23
⋮	⋮	⋮	⋮	⋱	⋮	⋮
19	20	21	22	⋯	37	38
20	21	22	23	⋯	38	39

Figure A.4: Number of Cubes on Each Square of a 20 × 20 Chessboard (A)

Note: The figure shows the number of cubes on each square of a chessboard, starting with one in the northwest corner and stepping up one each time you step south or east.

chessboard)[30]

$$
\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n}(i+j-1) &= \sum_{i=1}^{n}\sum_{j=1}^{n}[(i-1)+j] \\
&= \sum_{i=1}^{n}\left[n(i-1)+\frac{n(n+1)}{2}\right] \\
&= \left[n\left(\frac{n(n+1)}{2}-n\right)+n\left(\frac{n(n+1)}{2}\right)\right] \\
&= n\left(\frac{n^2+n-2n+n^2+n}{2}\right) \\
&= n^3.
\end{aligned}
$$

The answer n^3 is extremely neat and tidy. In the special case where $n = 20$, there are $20^3 = 8,000$ cubes on the board.

SECOND SOLUTION

With an answer this neat, there must be a non-algebraic solution. Imagine the 20 × 20 chessboard in front of you, with the stacks of cubes on it as in figure A.4. Now slice through the cubes horizontally at height 20 units. The cubes above the slice all lie in the southeast lower-triangular section below the non-leading diagonal. Now flip the above-the-slice cubes across the diagonal from southeast to northwest. They will fill the lower stacks to a height of 20 units. You now have a solid cube, and the result follows immediately.

Answer 1.35: The naive strategy is to run directly away from the dog toward the edge of the field. However, at speed v, it takes you $\frac{R}{v}$ units of time to get to the perimeter, while it takes the dog only $\frac{1}{2}\frac{2\pi R}{4v} = \frac{\pi R}{4v} \approx \frac{3}{4} \times \frac{R}{v}$ units of time to get there—so he will meet you and eat you. You somehow need to get further from him and closer to the fence before you make a run for it.

[30] I make use of the property that $\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$ (see Question 1.2).

Suppose you behave somewhat like the dog. Step away from the centre of the circle until you are at a radius of $\frac{R}{4}$. Now constrain yourself to running circuits around that radius. It takes you $\frac{\pi R}{4v}$ units of time to run half-way around this circle. The dog can also run half-way around the field in the same time. That is, at this radius, you and the dog are perfectly matched in your abilities to run around in circles.

Now suppose you step slightly closer to the centre of the circle. Let us say you now start running around in circles of radius $\frac{R}{4} - \epsilon$, for some small ϵ. In this case, you have a slight advantage over the dog: you can run around your circle in slightly less time than he can run around his. As you run, the dog tries to track you. However, you are gaining a little on the dog with every circuit. Eventually, you will be at the "top" of your circle, when he is at the "bottom" of his. Now it is time to make a run for it. You only have to travel a distance of $R - \left(\frac{R}{4} - \epsilon\right) = \frac{3}{4}R + \epsilon$. The dog has to travel a distance πR to meet you. You can outrun him as long as the time it takes you at speed v is less than the time it takes him at $4v$, that is

$$
\begin{aligned}
\frac{\frac{3}{4}R+\epsilon}{v} &< \frac{\pi R}{4v} \\
\Leftrightarrow \frac{3}{4}R+\epsilon &< \frac{\pi R}{4} \\
\Leftrightarrow 3R+4\epsilon &< \pi R \\
\Leftrightarrow \epsilon &< \frac{(\pi-3)R}{4}.
\end{aligned}
$$

It follows that if you choose an ϵ such that $0 < \epsilon < \frac{(\pi-3)R}{4}$, then you can run in a circle of radius $\frac{R}{4} - \epsilon$ until you are as far from the dog as possible and then you can escape by running away from him.

Answer 1.36: There are two methods. The first method assumes a known probability result (this may be sufficient for you); the second method subsumes the first by proving the aforementioned probability result before proceeding.

FIRST SOLUTION

The integral is immediately recognized as a simple transformation of an integral over the entire domain of a Normally distributed random variable.

The Standard Normal distribution has probability density function $f(u) \equiv \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$, for $-\infty < u < +\infty$. Integrating over the entire domain must produce total probability mass of unity:

$$
\int_{-\infty}^{+\infty} f(u)du = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du = 1
$$

If we substitute in $x = \frac{1}{\sqrt{2}}u$ (to make the integral look like the one we seek), then $dx = \frac{1}{\sqrt{2}}du$, and we get

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-x^2} dx = 1.$$

Multiply both sides by $\sqrt{\pi}$, and the result follows immediately.

SECOND SOLUTION

Let $I = \int_{-\infty}^{+\infty} e^{-x^2} dx$. Then squaring gives the following:

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{+\infty} e^{-x^2} dx\right) \cdot \left(\int_{-\infty}^{+\infty} e^{-y^2} dy\right) \\ &= \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} e^{-(x^2+y^2)} dy dx \\ &\stackrel{\text{see below}}{=} \int_{\theta=0}^{2\pi} \int_{r=0}^{+\infty} e^{-r^2} r dr d\theta \\ &= \int_{\theta=0}^{2\pi} \left[-\frac{1}{2}e^{-r^2}\right]_{r=0}^{\infty} d\theta \\ &= \frac{1}{2}\int_{\theta=0}^{2\pi} d\theta = \pi. \end{aligned}$$

Thus, $I = \sqrt{\pi}$, as required.

The above basis change from Cartesian coordinates to polar coordinates uses the transformation $x = r\cos\theta$, and $y = r\sin\theta$. This implies that $x^2 + y^2 = r^2$. However, there is more to it than this. You also need to know the general result

$$\int_x \int_y f(x,y) dx dy = \int_\theta \int_r f(x(r,\theta), y(r,\theta)) r dr d\theta.$$

The "r" in the integrand on the right-hand side is the "Jacobian" of the transformation from Cartesian to polar coordinates. The Jacobian, J, is just the determinant of the matrix of partial derivatives of the transformation. That is

$$\begin{aligned} \int_x \int_y f(x,y) dx dy &= \int_\theta \int_r f(x(r,\theta), y(r,\theta)) J dr d\theta, \text{ where} \\ J \equiv \frac{\partial(x,y)}{\partial(r,\theta)} &\equiv \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} \\ &= \begin{vmatrix} \frac{\partial[r\cos\theta]}{\partial r} & \frac{\partial[r\cos\theta]}{\partial \theta} \\ \frac{\partial[r\sin\theta]}{\partial r} & \frac{\partial[r\sin\theta]}{\partial \theta} \end{vmatrix} \\ &= \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} \\ &= r\cos^2\theta + r\sin^2\theta \\ &= r(\cos^2\theta + \sin^2\theta) = r. \end{aligned}$$

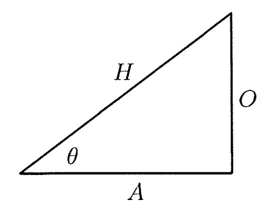
For more on Jacobians, determinants, and transformations, consult DeGroot (1989, pp162–166) and Anton (1988, pp1068–1069).

Answer 1.37: Of all the simple trigonometric functions that you might have been asked to integrate, $\int \sec\theta d\theta$ has arguably the most complicated answer.

Perhaps it is useful to review quickly the definitions of these trigonometry functions. Consider a right-angle triangle (see table A.3). Let θ be one of the acute angles (i.e., one of the two angles of less than 90 degrees). Let the lengths of the sides of the triangle be denoted by "A" (the side adjacent to the angle θ), "O" (the side opposite to the angle θ), and "H" (the hypotenuse—opposite the right angle), then the elementary trigonometric functions may be defined as in table A.3.[31]

Table A.3: Trigonometric Functions: Definitions

$\sin\theta = \frac{O}{H}$	$\text{cosec}\,\theta = \frac{H}{O} = \frac{1}{\sin\theta}$
$\cos\theta = \frac{A}{H}$	$\sec\theta = \frac{H}{A} = \frac{1}{\cos\theta}$
$\tan\theta = \frac{O}{A} = \frac{\sin\theta}{\cos\theta}$	$\cot\theta = \frac{A}{O} = \frac{1}{\tan\theta} = \frac{\cos\theta}{\sin\theta}$



Mnemonic: "SOH-CAH-TOA"

Note: These definitions are for the triangle illustrated. The sides are of length A (adjacent to the angle θ), O (opposite to the angle θ), and H (the hypotenuse).

For the particular problem given, $\int \sec\theta d\theta$, we see in table A.4 that the answer is $\ln|\sec\theta + \tan\theta|$ (up to an arbitrary constant of integration), which is readily verified via differentiation.

Answer 1.38: The sum $\sum_{n=1}^{\infty} e^{-\sqrt{n}}$ takes the form $\sum_{n=1}^{\infty} a_n$, where $a_n \equiv e^{-\sqrt{n}}$. There is a whole host of tests for the convergence of such sums. Before we look at these, a short review of the terminology is in order.

A "sequence" is a set of numbers a_1, a_2, a_3, ... indexed in a particular order corresponding to the natural numbers. We may denote the sequence as "$\{a_n\}$." Each number, a_n, in the sequence is a "term." The "limit of a sequence" exists

[31]The trigonometric functions' names are short for sine, cosine, tangent, cotangent, secant, and cosecant.

Table A.4: Trigonometric Functions: Calculus

$\int f(x)dx$	$f(x)$	$f'(x)$
$-\cos\theta$	$\sin\theta$	$\cos\theta$
$\sin\theta$	$\cos\theta$	$-\sin\theta$
$\ln\lvert\sec\theta\rvert$	$\tan\theta$	$\sec^2\theta$
$\ln\lvert\sin\theta\rvert$	$\cot\theta$	$-\mathrm{cosec}^2\,\theta$
$\ln\lvert\sec\theta+\tan\theta\rvert$	$\sec\theta$	$\sec\theta\tan\theta$
$\ln\lvert\tan\frac{1}{2}\theta\rvert$	$\mathrm{cosec}\,\theta$	$-\mathrm{cosec}\,\theta\cot\theta$

Note: The middle column gives a trigonometric function. The columns to the left and right give the integral of the function (ignoring arbitrary constant), and the derivative of the function, respectively.

and is equal to $l < \infty$ if the numbers a_n get closer and closer to l as n gets larger. That is, $\lim_{n\to\infty} a_n = l$. If such a limit exists, then the sequence is said to "converge" to that limit, and the limit is unique. If a sequence does not converge, then it "diverges." There is no mention of additivity here: A sequence is a succession, not a sum.

A "series" is formed from a sequence via partial sums. Let $S_1 = a_1$, $S_2 = a_1 + a_2$, $S_3 = a_1 + a_2 + a_3$, and so on, so that $S_n = \sum_{i=1}^{n} a_i$ is the nth "partial sum" of the sequence $\{a_n\}$. Then the sum $\sum_{n=1}^{\infty} a_n$ is referred to as an "infinite series." The infinite series $\sum_{n=1}^{\infty} a_n$ is said to be "convergent" if the sequence of its partial sums $\{S_n\}$ is convergent.

A necessary (but not sufficient) condition for convergence of an infinite series $\{a_n\}$ is that $a_n \to 0$ as $n \to \infty$. In our case, $a_n = e^{-\sqrt{n}} \to 0$, so we cannot reject convergence.

The first (of several) formal tests that comes to mind is **The Ratio Test** (for series with positive terms only):

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} \begin{cases} < 1, & \Rightarrow \sum a_n \text{ converges,} \\ > 1, & \Rightarrow \sum a_n \text{ diverges,} \\ = 1, & \Rightarrow \text{the test fails.} \end{cases}$$

In our case, the test fails because

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to\infty} \frac{e^{\sqrt{n}}}{e^{\sqrt{n+1}}} = 1.$$

The next formal test for convergence that comes to mind is **The nth Root Test** (for series with positive terms only):

$$\lim_{n\to\infty} \sqrt[n]{a_n} \begin{cases} < 1, & \Rightarrow \sum a_n \text{ converges,} \\ > 1, & \Rightarrow \sum a_n \text{ diverges,} \\ = 1, & \Rightarrow \text{the test fails.} \end{cases}$$

In our case, the test fails because

$$\begin{aligned} \lim_{n\to\infty} \sqrt[n]{a_n} &= \lim_{n\to\infty} \left(e^{-\sqrt{n}}\right)^{\frac{1}{n}} \\ &= \lim_{n\to\infty} e^{-n^{\frac{1}{2}}\times n^{-1}} \\ &= \lim_{n\to\infty} e^{-n^{-\frac{1}{2}}} \\ &= \lim_{n\to\infty} \frac{1}{e^{\frac{1}{\sqrt{n}}}} \\ &= 1. \end{aligned}$$

When the two tests above fail, we head for **Raabe's Test** (for series with positive terms only):

$$\lim_{n\to\infty} \left[n\left(1-\frac{a_{n+1}}{a_n}\right)\right] \begin{cases} > 1, & \Rightarrow \sum a_n \text{ converges,} \\ < 1, & \Rightarrow \sum a_n \text{ diverges,} \\ = 1, & \Rightarrow \text{the test fails.} \end{cases}$$

In our case, the test indicates convergence because

$$\lim_{n\to\infty} \left[n\left(1-\frac{a_{n+1}}{a_n}\right)\right] = \lim_{n\to\infty} \left[n\left(1-e^{\sqrt{n}-\sqrt{n+1}}\right)\right] > 1.$$

However, the algebraic proof that that last limit exceeds one is by no means elegant. Instead of proving it, I present another test of convergence that is both elegant and conclusive.

Story: 1. She wore a Walkman and said she could listen to me and the music at the same time. 2. Balding candidate abruptly excused himself. Returned to office a few minutes later, wearing a hairpiece.

Interview Horror Stories from Recruiters
Reprinted by kind permission of *MBA Style Magazine*

The Integral Test applies to a series $\sum a_n$ of positive terms. Let $A(x)$ denote the function of x obtained by replacing n in a_n by x. Then if $A(x)$ is decreasing and continuous for $x \geq 1$,

$$\sum_{n=1}^{\infty} a_n \quad \text{and} \quad \int_{x=1}^{+\infty} A(x)dx$$

either both converge, or both diverge (Anton [1988, p623]).

In our case, $a_n = e^{-\sqrt{n}}$. To test for convergence of $\sum_{n=1}^{\infty} a_n$, we may look at convergence of $\int_{x=1}^{+\infty} A(x)$, where $A(x) \equiv e^{-\sqrt{x}}$ is seen to be both decreasing and continuous for $x \geq 1$.

We need $\int e^{-\sqrt{x}}dx$. My first guess for this integral is

$$\int e^{-\sqrt{x}}dx = e^{-\sqrt{x}} + x^{\frac{1}{2}}e^{-\sqrt{x}} = (1+\sqrt{x})e^{-\sqrt{x}}.$$

However, differentiation shows that I am out by a factor of -2. If you cannot guess this directly, you need some practice with integration by parts. We get the following integral:

$$\begin{aligned}
\int_{x=1}^{+\infty} e^{-\sqrt{x}}dx &= -2(1+\sqrt{x})e^{-\sqrt{x}}\Big|_1^{\infty} \\
&= 2(1+\sqrt{x})e^{-\sqrt{x}}\Big|_{\infty}^{1} \\
&= 2(1+\sqrt{x})e^{-\sqrt{x}}\Big|_1 - 2(1+\sqrt{x})e^{-\sqrt{x}}\Big|_{\infty} \\
&= \frac{4}{e} - 2\lim_{x\to\infty}\frac{(1+\sqrt{x})}{e^{\sqrt{x}}} \\
&= \frac{4}{e} - 2\lim_{u\to\infty}\frac{(1+u)}{e^{u}} = \frac{4}{e},
\end{aligned}$$

because $\sqrt{x} \to \infty$ if and only if $x \to \infty$, and $\lim_{u\to\infty}\frac{(1+u)}{e^u} = 0$ is well known. It follows that the series is convergent! Incidentally, the limit of the series $\sum_{n=1}^{+\infty} e^{-\sqrt{n}}$ is only slightly below $\frac{4}{e}$.

The **Comparison Test** says that if there exists $N < \infty$ such that $0 \leq a_n \leq b_n$ for $n \geq N$, and if $\sum_{n=1}^{+\infty} b_n$ is convergent, then so too is $\sum_{n=1}^{+\infty} a_n$. This also works in reverse: If $\sum_{n=1}^{+\infty} a_n$ is divergent, so too is $\sum_{n=1}^{+\infty} b_n$. This test requires that you construct b_n. In our case, if $n \geq 75$, then $a_n = e^{-\sqrt{n}} < \frac{1}{n^2} = b_n$, and $\sum_{n=1}^{+\infty}\frac{1}{n^2}$ is known to converge! In fact, $\sum_{n=1}^{+\infty}\frac{1}{n^2} = \frac{\pi^2}{6}$ (Spiegel [1968, p108]).

It is worth noting that if $\sum |a_n|$ converges, then $\sum a_n$ converges also. The former is referred to as "absolute convergence." Thus, absolute convergence of an infinite series is sufficient for convergence. Absolute convergence is not, however, a necessary condition for convergence. A series that is convergent, but not absolutely convergent, is said to be "conditionally convergent."

A final convergence test we might have tried is **Gauss' Test** (for series with positive terms only): If $\frac{a_{n+1}}{a_n} = 1 - \frac{\mathcal{L}}{n} + \frac{b_n}{n^2}$, where there exists an $M > 0$, and an N such that $|b_n| < M$ for all $n > N$, then the series $\sum_{n=1}^{\infty} a_n$ is

1. convergent if $\mathcal{L} > 1$, and
2. divergent or conditionally convergent if $\mathcal{L} \leq 1$.

For more information on tests of convergence of series, look to your favorite calculus book. Most of the above-mentioned tests should appear if the book is worthwhile.

Answer 1.39: You win if you can place the last coin on the table and leave no space for me to place a further coin. A necessary condition is that the table be radially symmetric. That is, there must exist a central point on the table (at its centre of mass if the table is of uniform density and thickness) such that any line drawn upon the table passing through this central point is evenly bisected at this central point. Simple examples are a square, an ellipse, a rectangle, a circular disc, etcetera.

You should play first and place your first quarter at the centre of the (round, square, ...) table. You should make subsequent moves by imitating me: Place your quarter in the mirror image of my position when viewed looking through the central point. This ensures victory because if I can still place a coin on the table, then so can you.[32]

Although radial symmetry is necessary, it is not sufficient. The strategy does not necessarily work if the table is a regular shape but not a simply-connected one; for example, an annulus.[33] If the table is an annulus and the hole in the middle is bigger than a quarter, then the only change to your winning strategy is that you should let me go first.

Answer 1.40: The numbers used and the situation described may differ from question to question, but the general solution technique is always the same. Factor the product into all possible triplets: (1,2,18), (1,3,12), (1,4,9), (1,6,6), (2,2,9), (2,3,6), and (3,3,4). Which one is it? Well, Mary knows the sum, and these potential triplets sum to 21,16,14,13,13,11, and 10, respectively. Knowing the sum was not sufficient for Mary to pin down the triplet, so it must be a triplet with a non-unique sum: 13 in this case. This cuts down the candidates to (1,6,6) and (2,2,9). John says the eldest is dyslexic, so there must be an eldest (ignoring rubbish answers like one twin is 20 minutes older than the other). That just leaves (2,2,9).

Answer 1.41: The sums in table A.5 are well known (the first is discussed extensively in the answer to Question 1.2).[34] You should certainly know the first sum by heart, and you should note that the third is the first squared (Grahame Bennett has given me a very elegant geometrical proof of this). My first solution uses a sensible guess plus induction. My second solution is similar, but requires that you notice or already know a special result.

[32] I thank Tim Hoel and Victor H. Lin for this elegant solution technique.

[33] An "annulus" is a disc with a hole in the centre—like a musical compact disc, for example. An annulus is path connected (any two points may be joined by a line), and is therefore connected (it cannot be split into two non-empty non-intersecting open sets), but it is not simply connected (which requires path-connectivity and that any loop may be shrunk continuously within the set).

[34] The fourth-order result is not well known: $\sum_{k=1}^{n} k^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}$ (Spiegel [1968, p108]).

Table A.5: Sums of k, k^2, and k^3

$\sum_{k=1}^{n} k$	$\frac{n(n+1)}{2}$
$\sum_{k=1}^{n} k^2$	$\frac{n(n+1)(2n+1)}{6}$
$\sum_{k=1}^{n} k^3$	$\frac{n^2(n+1)^2}{4}$

FIRST SOLUTION[35]
Given $\sum_{k=1}^{n} k = n(n+1)/2$, let us assume $\sum_{k=1}^{n} k^i$ equals an $(i+1)^{th}$-order polynomial $f^{(i)}(n)$. In the case $i = 2$ (i.e., trying to find $\sum_{k=1}^{n} k^2$), let this polynomial be $f^{(2)}(n) = an^3 + bn^2 + cn + d$. We can calculate $f^{(2)}(n)$ for $n = 1$–4, and set up the system equations A.1.

$$\begin{array}{rcrcrcrcrcr} f^{(2)}(1) & = & a & + & b & + & c & + & d & = & 1 \\ f^{(2)}(2) & = & 8a & + & 4b & + & 2c & + & d & = & 5 \\ f^{(2)}(3) & = & 27a & + & 9b & + & 3c & + & d & = & 14 \\ f^{(2)}(4) & = & 64a & + & 16b & + & 4c & + & d & = & 30 \end{array} \tag{A.1}$$

Standard row-reduction techniques soon yield $a = 1/3$, $b = 1/2$, $c = 1/6$, and $d = 0$. Thus, $f^{(2)}(n) = \frac{2n^3+3n^2+n}{6} = \frac{n(n+1)(2n+1)}{6}$. Having obtained a formula that works for $n = 1$–4, we must now prove, by induction, that if it works for n, then it works for $n+1$. That is, we show that $f(n) + (n+1)^2 = f(n+1)$:

$$\begin{aligned} f(n) + (n+1)^2 &= \frac{2n^3 + 3n^2 + n}{6} + (n^2 + 2n + 1) \\ &= \frac{2n^3 + 9n^2 + 13n + 6}{6} \\ &= \frac{(n+1)(n+2)(2n+3)}{6} \\ &= \frac{[(n+1)][(n+1)+1][2(n+1)+1]}{6} \\ &= f(n+1), \text{ as required.} \end{aligned} \tag{A.2}$$

The case $f^{(3)}(n) = \sum_{k=1}^{n} k^3$ may be proved in exactly the same fashion, and I leave it as an exercise.

SECOND SOLUTION
If you notice that $\sum_{k=1}^{n} 1 = n$ and $\sum_{k=1}^{n} k = n(n+1)/2$, you might deduce

[35] I thank Vince Moshkevich for suggesting this technique. Any errors are mine.

the following pattern:[36]

$$\begin{aligned} \sum_{k=1}^{n} 1 &= n \\ \sum_{k=1}^{n} k &= \frac{n(n+1)}{2} \\ \sum_{k=1}^{n} k(k+1) &= \frac{n(n+1)(n+2)}{3} \\ \sum_{k=1}^{n} k(k+1)(k+2) &= \frac{n(n+1)(n+2)(n+3)}{4} \\ &\vdots \end{aligned}$$

These can each be proved easily using mathematical induction. For example, the third equality immediately above is easily proved true when $n = 1$ (both sides equal 2). To prove this third equality in general, we now need only show that increasing n by one on each side of the equality has the same incremental effect on both sides. That is, we need only show that the right-hand side evaluated at $n = (N+1)$ less the right-hand side evaluated at $n = N$ gives what would be the $(N+1)^{st}$ term in the summation on the left-hand side:

$$\begin{aligned} & \frac{(N+1)(N+2)(N+3)}{3} - \frac{N(N+1)(N+2)}{3} \\ &= \frac{(N+1)(N+2)}{3}((N+3) - N) \\ &= \frac{(N+1)(N+2)}{3}(3) \\ &= (N+1)(N+2) \\ &= k(k+1)\Big|_{k=N+1} \quad \text{QED.} \end{aligned}$$

The results we are interested in follow quite easily now because, for example, $\sum_{k=1}^{n} k^2 = \sum_{k=1}^{n} k(k+1) - \sum_{k=1}^{n} k$, and we have expressions for both the latter summations.

Answer 1.42: You know one of the eight balls is heavy. Compare one group of three to another group of three. You need only one more weighing—for a total of two weighings.

Answer 1.43: Deciphering the optimal strategy is analogous to locating an optimal stock price exercise boundary for an American-style option. Calculating the expected payoff to the game, assuming the optimal strategy, is analogous to valuing an American-style option. Like valuing an American option, you

[36] I thank David Maslen for suggesting this technique. Any errors are mine.

have to work backward through a decision tree, calculating the expected payoffs to proceeding versus stopping at each node. For two, four, six, and eight cards, the expected payoff to the game is $\$\frac{1}{2}$, $\$\frac{2}{3}$, $\$\frac{17}{20}$, and \$1, respectively, when following the optimal strategy. The two-card game decision tree is a sub-tree of the four-card game decision tree, so later results can be appended to earlier ones. Stop reading here and try to replicate these numbers.

Let R and B denote the number of red and black cards, respectively, when you begin play ($R = B = 26$ in our case). Let r and b denote the number of red and black cards remaining in the deck at some intermediate stage of the game when you are trying to decide whether to take another card. You get $+1$ for each red card drawn and -1 for each black card drawn, so your current accumulated score is the number of reds drawn so far less the number of blacks drawn so far: $(R - r) - (B - b)$. The expected value of the game $V(r, b)$ is the current accumulated score plus the additional expected value, if any, remaining in the deck, assuming optimal play. With r and b cards remaining, denote this additional expected value as $E(r, b)$. Thus, the value of the game is $V(r, b) = (R - r) - (B - b) + E(r, b)$. Simple logic dictates that $E(r, b)$ is defined recursively as follows:[37]

$$E(r,b) = \begin{cases} 0, \text{ if } r = 0 \\ r, \text{ if } b = 0 \\ \max\left\{0, \frac{r}{r+b}[1 + E(r-1,b)] + \frac{b}{r+b}[-1 + E(r,b-1)]\right\}, \\ \text{otherwise.} \end{cases}$$

Table A.6 gives all the necessary information for you to solve the easier game in which there are four cards of each color. It is worth noting that the recursive definition of $E(r, b)$, when seen in action in table A.6, produces a complicated Pascal's Triangle type of calculation when working from the lower right to the upper left.[38] That is, $E(r, b)$ in each cell depends on $E(r, b)$ in the cells immediately to the right and below. As mentioned previously, in the two-, four-, six-, and eight-card games, the expected payoffs are $\$\frac{1}{2}$, $\$\frac{2}{3}$, $\$\frac{17}{20}$, and \$1, respectively, and these are visible on the leading diagonal in table A.6.

If the additional remaining value in the deck when playing optimally is zero [i.e., $E(r, b) = 0$], you are not indifferent about continuing. Rather, you want to quit immediately because in every case except one, $E(r, b) = 0$ implies $\frac{r}{r+b}[1 + E(r - 1, b)] + \frac{b}{r+b}[-1 + E(r, b - 1)]$ is negative, and that the "max" function is being used in the recursive definition of $E(r, b)$. The only exception is when $(r, b) = (1, 2)$, and even then $\frac{r}{r+b}[1+E(r-1, b)]+\frac{b}{r+b}[-1+E(r, b-1)]$ is zero and a risk-averse player would quit. When playing optimally, the last

[37] I thank Paul Turner for solving this problem when it was posted as a challenge question on my web site. Any errors are mine.

[38] Pascal's Triangle has the following rows: [1], [1 1],[1 2 1], [1 3 3 1], [1 4 6 4 1], and so on. Apart from the 1's, each item is the sum of the two items above. The $(n + 1)^{st}$ row gives the coefficients in the polynomial expansion of $(a + b)^n$.

Table A.6: The Red/Black Card Game

0	(4,4)	1	(3,4)	2	(2,4)	3	(1,4)	4	(0,4)
1	$(\frac{4}{8}, \frac{4}{8})$	$\frac{12}{35}$	$(\frac{3}{7}, \frac{4}{7})$	0	$(\frac{2}{6}, \frac{4}{6})$	0	$(\frac{1}{5}, \frac{4}{5})$	0	$(\frac{0}{4}, \frac{4}{4})$
1	Y	$1\frac{12}{35}$	Y	2	$N\bullet$	3	NN	4	NN
-1	(4,3)	0	(3,3)	1	(2,3)	2	(1,3)	3	(0,3)
$1\frac{23}{35}$	$(\frac{4}{7}, \frac{3}{7})$	$\frac{17}{20}$	$(\frac{3}{6}, \frac{3}{6})$	$\frac{1}{5}$	$(\frac{2}{5}, \frac{3}{5})$	0	$(\frac{1}{4}, \frac{3}{4})$	0	$(\frac{0}{3}, \frac{3}{3})$
$\frac{23}{35}$	Y	$\frac{17}{20}$	Y	$\frac{6}{5}$	Y	2	$N\bullet$	3	NN
-2	(4,2)	-1	(3,2)	0	(2,2)	1	(1,2)	2	(0,2)
$2\frac{2}{5}$	$(\frac{4}{6}, \frac{2}{6})$	$1\frac{1}{2}$	$(\frac{3}{5}, \frac{2}{5})$	$\frac{2}{3}$	$(\frac{2}{4}, \frac{2}{4})$	0	$(\frac{1}{3}, \frac{2}{3})$	0	$(\frac{0}{2}, \frac{2}{2})$
$\frac{2}{5}$	Y	$\frac{1}{2}$	Y	$\frac{2}{3}$	Y	1	$N\bullet$	2	NN
-3	(4,1)	-2	(3,1)	-1	(2,1)	0	(1,1)	1	(0,1)
$3\frac{1}{5}$	$(\frac{4}{5}, \frac{1}{5})$	$2\frac{1}{4}$	$(\frac{3}{4}, \frac{1}{4})$	$1\frac{1}{3}$	$(\frac{2}{3}, \frac{1}{3})$	$\frac{1}{2}$	$(\frac{1}{2}, \frac{1}{2})$	0	$(\frac{0}{1}, \frac{1}{1})$
$\frac{1}{5}$	Y	$\frac{1}{4}$	Y	$\frac{1}{3}$	Y	$\frac{1}{2}$	Y	1	$N\bullet$
-4	(4,0)	-3	(3,0)	-2	(2,0)	-1	(1,0)	0	(0,0)
4	$(\frac{4}{4}, \frac{0}{4})$	3	$(\frac{3}{3}, \frac{0}{3})$	2	$(\frac{2}{2}, \frac{0}{2})$	1	$(\frac{1}{1}, \frac{0}{1})$	0	$(0, 0)$
0	Y	0	Y	0	Y	0	Y	0	$N\bullet$

Note: Each cell is laid out as

Accum. Score	(r, b)
$E(r, b)$	(p_{red}, p_{black})
$V(r, b)$	$Y, N,$ or NN

, where r and b are the number of red and black cards remaining in the deck, "Accum. Score" is the accumulated score so far (i.e., $(R - r) - (B - b)$, where $R = B = 4$ in this case); p_{red} and p_{black} are the probabilities that the next card drawn will be red or black, respectively; $E(r, b)$ is the expected additional value remaining in the deck assuming optimal play; $V(r, b)$ is the expected payoff of the game, assuming you start in the top left cell (it is the sum of the two items above it); "Y" means *yes* you should continue playing; "$N\bullet$" means *no* you should halt (the bullet helps your eye see the boundary), and "NN" means *no* you should halt, but you should also note that it is *not* possible to get to this cell if you start with an even number of each color card and play optimally.

card drawn is always red. That is, you never pick a black card and then quit. The optimal score to quit at cannot be negative because you always have the safety net of a zero payoff for sure if you draw every card. The optimal score to quit at is a non-increasing step function of the number of black cards drawn (drawing red cards has no effect on the optimal score to quit at). Drawing black cards can lower the optimal score to quit at. In the eight-card game, the optimal stopping rule is: If you have turned over zero or one black card, then quit if you can get to a score of 2 without seeing another black card; if you have turned over two or three black cards, then quit if you can get to a score of 1 without seeing another black card; if you have turned over four black cards, then the best you can do is draw every card and get a payoff of 0. In the $2n$ card game with n red cards and n black cards, the expected payoffs are shown in table A.7.[39]

Table A.7: E(Payoff) in Red/Black Card Games ($2n$ cards, n red, n black)

$2n$	$r=b=n$	$V(r,b)$ (ratio)	$V(r,b)$ (decimal)
2	1	$\frac{1}{2}$	0.500000000000
4	2	$\frac{2}{3}$	0.666666666667
6	3	$\frac{17}{20}$	0.850000000000
8	4	$\frac{1}{1}$	1.000000000000
10	5	$\frac{47}{42}$	1.119047619048
12	6	$\frac{284}{231}$	1.229437229437
14	7	$\frac{4{,}583}{3{,}432}$	1.335372960373
⋮	⋮	⋮	⋮
52	26	$\frac{41{,}984{,}711{,}742{,}427}{15{,}997{,}372{,}030{,}584}$	2.624475548994

Note: These expected payoffs are derived using the same rules used in the eight-card game. I have included the ratio form of the expected payoff in case anyone spots a simple pattern.

Answer 1.44: No, definitely not. You cannot tile the 62 squares with the dominoes. If you cannot see why, then go back and think again. This one is too good to waste by peeking at the answers—stop reading here and try again before reading further.

Each domino covers two squares that are side-by-side on the board. Each of these pairs of squares consists of a black and a white. As you place the dominoes, you cover the same number of black squares as white ones. However, the two squares that are off limits are the same color (opposing corners on a chessboard must be the same color). Thus, the number of white squares to be covered is not the same as the number of black, and the dominoes cannot cover them all.

Naoki Sato has supplied an answer to his follow up question. Imagine a closed path on the chessboard that passes through every square exactly once (moving horizontally and vertically, eventually returning to the original square). The two "X"s, unless adjacent, divide this path into two sections. Since one "X" is on black, and one is on white, the two sections each cover an even number of squares. They may thus be tiled using the dominoes. If the two "X"s are adjacent, the solution is obvious.

[39] I thank David Maslen for the final ratio in the table. Any errors are mine.

Answer 1.45: A prime number has no factors other than itself and 1. Thus, 4 is not prime because it has factors: (1,4), and (2,2). Drawing a number line might be a good way to explain this to an interviewer, but I will just use words.

1. A prime p bigger than 2 cannot be an integer multiple of 2, else it would not be prime. Thus, a prime bigger than 2 must be odd. Thus, $p-1$ is even. Thus, $p-1=2n$ for some positive integer n. Thus, $p=2n+1$.
2. A prime p bigger than 3 cannot be an integer multiple of 3, else it would not be prime. However, draw a number line and it must be that either $p-1$ or $p+1$ (but not both) is a multiple of 3. That is, p is 1 away from a multiple of 3, but we do not know in which direction. Thus, $p\pm1=3m$ for some positive integer m, where $\pm$ means exactly one of $+$ or $-$, but not both. Thus, $p=3m\pm1$.
3. The question asks about p^2-1. From #1 we see that $p^2-1=4n^2+4n=4n(n+1)$. One of n or $n+1$ must be even, and with that 4 there, we see that p^2-1 contains a factor of 8 (i.e., $2\times2\times2$).
4. From #2, we see that $p^2-1=9m^2\pm6m=3m(m\pm2)$. Thus, p^2-1 contains 3 as a factor.
5. If we picture p^2-1 factored out into all possible numbers of smallest possible size, then the results from #3 and #4 cannot overlap. That is, p^2-1 contains factors of $2\times2\times2$ and 3; thus, p^2-1 is an integer multiple of 24.

Answer 1.46: Let B be your bid. Let S be the true value of the firm. The density function of S equals unity for $0\leq S\leq1$, and zero otherwise. Your payoff P is

$$P(S)=\begin{cases}2S-B, & \text{if } B>S\\ 0, & \text{otherwise.}\end{cases}$$

The maximum post-bid firm value is 2, so you should bid no more than 2. You want to maximize $E[P(S)]$ with respect to choice of B in the interval $[0,2]$. Your expected payoff is

$$\begin{aligned}E[P(S)] &= \int_{S=0}^{S=1} P(S)\cdot 1\cdot dS\\ &= \int_{S=0}^{S=\min(B,1)} (2S-B)dS\\ &= (S^2-BS)\Big|_{S=0}^{S=\min(B,1)}\\ &= \begin{cases}0, & \text{if } B\leq1\\ 1-B, & \text{if } B>1,\end{cases}\end{aligned}$$

so you should bid less than or equal to 1 and expect to break even.

Answer 1.47: What is going to happen if you light both ends simultaneously? The two fizzing sparking flames are going to burn toward each other and meet. When they meet 60 seconds worth of fuse has been burnt in two sections that each took the same amount of time. How much time? It has to be exactly 30 seconds because they both took the same time, and these times add to 60 seconds. Of course, you have to bend the fuse so that you can light both ends simultaneously and when they meet it probably won't be in the centre of the fuse.

Answer 1.48: Light Fuse 1 at both ends and simultaneously light Fuse 2 at one end. As soon as Fuse 1 is burned out (i.e., after 30 seconds), light the other end of Fuse 2.

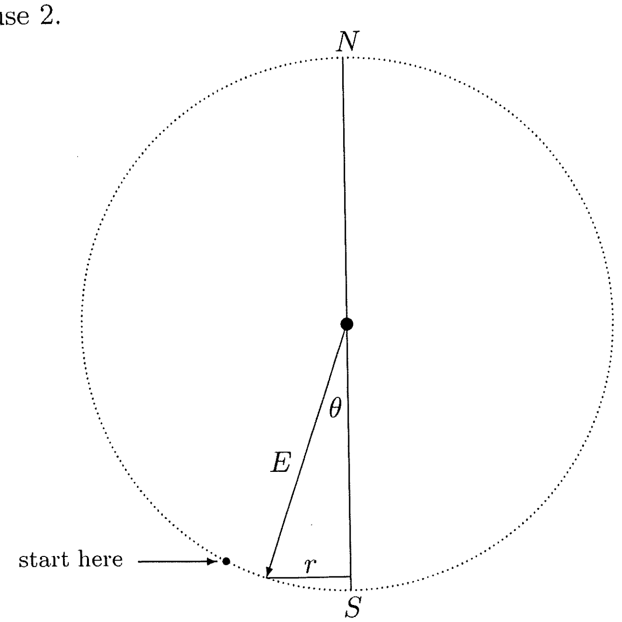


Figure A.5: S-E-N Problem: The Earth

Note: The Earth is a perfect sphere with radius E. You start your trek one mile north of a line of latitude having circumference $1/n$ miles, and radius r miles (so $2\pi r = 1/n$). You must start a distance of $1 + E \cdot \arcsin \frac{1}{2\pi n E}$ miles from the south pole—see Answer 1.49.

Answer 1.49: If your answer is "none" or "one," then go back and think again. There are, in fact, an uncountably infinite number of starting points that solve this problem.

First of all, you could start at the north pole. On the middle leg of your walk you would always be one mile south of the north pole, so the final leg would put you back where you started. Second, if you start at a point close to the south pole but one mile north of a line of latitude having circumference one

mile, then the middle leg of your walk begins and ends in the same spot; the final leg takes you back to your starting point. There are infinitely many such starting points on the line of latitude that is one mile north of the line of latitude having circumference one mile.

Similarly, if you start slightly further south, at a point one mile north of a line of latitude having circumference one-half mile, then the middle leg of your walk begins and ends in the same spot, and the final leg returns you to your starting point.

More generally, if you begin on a line of latitude one mile north of a line of latitude having circumference $1/n$ miles, then you will walk one mile south, circle the line of latitude n times, and return to your starting point.

In the latter case, how far is your starting point from the south pole? Well, assume the Earth is perfectly spherical, and let E be its radius. Let r be the radius of the line of latitude having circumference $1/n$ miles, so, $2\pi r = 1/n$. A simple sketch shows that the angle θ between the axis of the Earth, and a line drawn from the centre of the Earth to any point on the line of latitude having circumference $1/n$ miles, satisfies $\sin\theta = r/E$ (see figure A.5 and the trig' review on page 87). Thus, the arc length from the pole to this line of latitude is the fraction $\frac{\arcsin \frac{r}{E}}{2\pi}$ of the circumference of the Earth, $2\pi E$. That is, the arc length is $E \cdot \arcsin \frac{1}{2\pi n E}$ (using $r = 1/(2\pi n)$). You start one mile north of this, at a distance of $1 + E \cdot \arcsin \frac{1}{2\pi n E}$ miles from the south pole.

Answer 1.50: $100! = 100 \times 99 \times 98 \times \ldots 3 \times 2 \times 1$. Factor each number and count how many supply a 5. Combine the 5's with all the 2's going spare to get the 10's that give 0's at the end of 100!. The following supply a 5 (or two, as indicated): 5, 10, 15, 20, 25(2), 30, 35, 40, 45, 50(2), 55, 60, 65, 70, 75(2), 80, 85, 90, 95, 100(2). This gives the 24 zeroes at the end of 100!:[40]

```
                        933 26215
44394 41526 81699 23885 62667
00490 71596 82643 81621 46859
29638 95217 59999 32299 15608
94146 39761 56518 28625 36979
20827 22375 82511 85210 91686
40000 00000 00000 00000 00000
```

Answer 1.51: The king should take one coin from bag one, two coins from bag two, three coins from bag three, and so on, finishing with ten coins from bag ten. Place this collection on the weighing device, and look for the discrepancy from $\sum_{i=1}^{10} i$ ounces. If the actual weight is 0.40 ounces short, for example, then bag four is light, and collector four is the cheat.

[40]Type `vpa factorial(100) 158` in MATLAB; vpa is variable precision arithmetic.

Answer 1.52: Snap the bar into pieces that are one, two, and four parts long, respectively. On day one, give him one part. On day two, exchange your two parts for his one. On day three, give him back the one part. On day four, exchange four parts for his three. On day five, give him one more part. On day six, exchange your two parts for his one. On day seven, give him back the one part.

Answer 1.53: Let us attack the mirror problem in stages.

Your Perspective, No Rotations: Put your wristwatch on your left wrist and stand facing a mirror with your arms held out as though you are being crucified (it is a tough interview remember). Your reflected self's wristwatch-bearing arm is pointing the *same* direction as yours. Your wristwatch is to the left of your head, and your reflected self's wristwatch is also to *your* left of his or her head. *There has been no flipping of left for right.* Similarly, if your head is pointing up, then your reflected self's head is also pointing up, and *there has been no flipping of up for down.*

Perhaps this is clearer if you write a sentence on a transparent plastic sheet, and hold the sheet in front of your body, as though there is no mirror at all and you are simply reading what you have just written. Now look in the mirror. The reflection of your sheet in the mirror is *not* reversed. That is, the left-most word is still left most, the right-most word is still right most, and you can still read the reflected image from left to right.

Viewed from your perspective, everything about you that is left, right, up, or down is still left, right, up, or down, respectively, in your reflected image. There is thus *no* flipping of left for right or up for down. What *has* flipped is that if you are facing east, then your reflection is facing west. It does not matter for the sentence written on the transparent sheet, because it has no depth. It does matter for you, because your reflected nose is facing the opposite direction.

Your Perspective, Rotation of Yourself: If your interviewer suggests that there really is a flipping left for right of your reflected self, but not up for down, then this requires an implicit rotation of your perspective about a vertical axis, to place your right-handed self into the imagined boots of your reflected self who is left handed and standing on the other side of the mirror. To get a one-to-one mapping (so the boots fit), however, you still need to flip yourself left for right (without changing the direction in which you are facing) because your wristwatch is on your left wrist—the opposite of your reflected self. Had you instead rotated yourself about a horizontal axis, and then attempted to place yourself into the imagined boots of your reflected self, you would find your noses pointing the same direction and your wristwatches on the same sides, but your head would be between the feet of your reflected self, and to get into those imagined boots, you would still need to flip yourself up for down (without changing the direction in which you are facing).

The fact that neither a rotation about a horizontal nor a vertical axis suffices to place you into the imagined boots of your reflected self, serves to confirm my earlier assertion: There has not been a flipping of left for right, or up for down, but rather, a flipping in the direction of the depth. If your interviewer firmly believes that a mirror does flip left for right, then he or she is predisposed toward rotation about a vertical axis (something many of us do every day), and has not thought through the consequences of the attempted one-to-one mapping.

Answer 1.54: Yes, it can be done, in theory if not in practice. If you are stuck and looking for a hint, think about inverting a condom and covering it with another.

Let us label the condoms $C1$, and $C2$, and the men $M1$, $M2$, and $M3$. $M1$ wears $C1$ with $C2$ placed over it. $M2$ then uses $C2$, which is still clean inside. $M3$ then wears $C1$ inverted ($C1$'s outside, you will recall, was kept clean by $C2$), and places the twice-used $C2$ over it. Don't try this at home.

Answer 1.55: You will take a total of ten steps. Five of these steps will be east; five will be north. You only need to choose which five of the ten steps are east. There are $\binom{10}{5} = \frac{10!}{5!(10-5)!} = \frac{10\cdot 9\cdot 8\cdot 7\cdot 6}{5\cdot 4\cdot 3\cdot 2\cdot 1} = \frac{30,240}{120} = 252$ ways to make this choice.

Answer 1.56: Easier than it looks! If the bill is X, then with the tip it is $1.2X$ and $\frac{1.2}{6}$ is just one fifth (i.e., 0.20). So, all you have to do is multiply the quoted bill by 2 and move the decimal place! The bill was 132.67, so times 2 gives 265.34 and move the decimal place to get \$26.534 (you should be able to do that in your head). The key here is that $\frac{1.2}{6}$ multiplied by 132.67 has much more structure than dividing the multiple of 1.2 and 136.67 by 6.

Answer 1.57: We need to figure out whether the area of the large pizza is greater or smaller than the sum of the areas of the medium and small pizzas. The area of a circle is proportional to the square of the diameter. So, let L, M, and S be the diameters of the three pizzas, respectively. We need only take half of each pizza and lay the three halves on the table so that the corners touch and the three diameters form a triangle. If the angle in the corner where the small and medium pizzas touch is a right angle (check it using one corner of the pizza box!), then $L^2 = M^2 + S^2$ holds by Pythagoras' Theorem and the two orders are equally attractive. If the angle is larger than a right angle, then $L^2 > M^2 + S^2$ and the large pizza is the better deal. If the angle is smaller than a right angle, then $L^2 < M^2 + S^2$ and the small pizza plus medium pizza is the better deal. We could alternatively use a single slice from each pizza (with side length equal to the radius of the circle) and form a triangle of side lengths.

Answer 1.58: A sixth order polynomial has six roots. From complex analysis you may recall that the roots to this sort of equation are distributed evenly on a circle in the complex plane of radius equal to the positive real root. So, the roots are $z_k = 2\,e^{\frac{k2\pi i}{6}}$ for $k = 0, 1, 2, 3, 4, 5$. Note that $e^{\pi i} = -1$, so $2e^{\pi i} = -2$, which makes sense.

Answer 1.59: We will use the standard high school physics equations for linear motion with constant acceleration. There are, however, two very good reasons for having a sensible guess at the answer before doing any math. First, if our calculations give an answer that is wildly different from our guess, then we have some baseline figure for suspecting we may have made an error in the math. Second, if you start with a guess, it may be that your interviewer will hold up a hand and say "OK, fine, let's move to the next question." He or she may just have wanted to get an estimate out of you to see if you can estimate anything. They may not interested at all in whether you remember simple equations of motion. So, I would say "Well, I can estimate it using equations for linear motion with constant acceleration, but first let me guess that it is something like ...five seconds and 100 miles per hour" (or whatever *your* guess is).

Now, if we have to go on to the math, I would tell the interviewer that these equations of motion ignore air resistance. The penny may in fact reach "terminal velocity" (i.e., when drag from the air resistance produces an upward force that perfectly counters the force of gravity, and the penny stops accelerating). Even if it does not reach terminal velocity my estimate of speed will be an upper bound only because the air resistance will slow the penny down.

Now, to do the math, we need to know how high the Empire State Building is. From memory it is about 100 floors. Looking at the building I am sitting in, I guess that each floor is probably about four meters, so let us assume it is 400 meters high.

Now, let v be final velocity, u be initial velocity (i.e., zero), D be the 400m traveled, and t be the time taken. We can use $v^2 = u^2 + 2 \cdot g \cdot D$, where $g = 9.8m/s^2 \approx 10m/s^2$ is the acceleration due to gravity. Then $v^2 \approx 2 \cdot 10 \cdot 400 = 8{,}000m^2/s^2$, so $v \approx 90m/s$ (because $9^2 = 81$, so $90^2 = 8{,}100$). I keep track of units so that I can do a "dimensional analysis" (i.e., so that I can quickly confirm that the units of the RHS match the units of the LHS).

Then we can use $v = u + g \cdot t$ to deduce that $90 \approx 0 + 10 \cdot t$, so $t \approx 9s$. So, it takes nine seconds, and it is traveling at 90 m/s when it hits the ground. What is this in miles per hour (mph)? Well, we need to multiply by 60 and then again by 60 to get meters per hour, then divide by 1,609m/mile to get miles per hour. I don't even need the back of an envelope to get $90 \times 60 \times 60/1{,}609 = 324{,}000/1{,}609$, but that's very close to $320{,}000/1{,}600$ which is just 200. So, it takes nine seconds and hits the ground at 200 miles per hour. That's not wildly different from my initial guess, so I have no reason to suspect any error. I would now tell the interviewer that this is an upper bound on speed, and a lower bound on time taken.

Answer 1.60: You are asked to express the integral $f(x) = \int_{t=x}^{\infty} e^{-a\frac{t^2}{2}+bt}dt$ in

terms of $N(x)$. If you take that literally, then you will get stuck because that cannot be done for general x. Rather, we will aim to express the integral as a function of $N(g(x))$ for some $g(\cdot)$.

We are aiming for the integrand to take the functional form of the pdf of the Standard Normal: $n(u) \equiv \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$. So, we need to "complete the square" in the exponent, change variables, and be sure to remember the $\sqrt{2\pi}$ multiplier. Let us focus on completing the square in the exponent first; we need the minus one half multiplier, so we can pull that out and then add and subtract half the square of the coefficient of the linear term as follows:

$$
\begin{aligned}
-a\frac{t^2}{2} + bt &= -\frac{1}{2}\left(at^2 - 2bt\right) \\
&= -\frac{1}{2}\left[a\left(t^2 - 2\frac{b}{a}t\right)\right] \\
&= -\frac{1}{2}\left[a\left(t^2 - 2\frac{b}{a}t + \left(\frac{b}{a}\right)^2 - \left(\frac{b}{a}\right)^2\right)\right] \\
&= -\frac{1}{2}\left[a\left(\left(t - \frac{b}{a}\right)^2 - \left(\frac{b}{a}\right)^2\right)\right] \\
&= -\frac{1}{2}a\left(t - \frac{b}{a}\right)^2 + \frac{b^2}{2a}
\end{aligned}
$$

Now for the change of variables. Let $u = \sqrt{a}\left(t - \frac{b}{a}\right)$. Then $du = \sqrt{a}dt$, and $t = x \Rightarrow u = \sqrt{a}\left(x - \frac{b}{a}\right)$. So, making the change of variables and collecting terms, we get

$$
\begin{aligned}
f(x) &= \int_{t=x}^{\infty} e^{-a\frac{t^2}{2}+bt}dt \\
&= \frac{1}{\sqrt{a}}\int_{u=\sqrt{a}\left(x-\frac{b}{a}\right)}^{\infty} e^{-\frac{1}{2}u^2+\frac{b^2}{2a}}du \\
&= e^{\frac{b^2}{2a}}\sqrt{\frac{2\pi}{a}}\left[\frac{1}{\sqrt{2\pi}}\int_{u=\sqrt{a}\left(x-\frac{b}{a}\right)}^{\infty} e^{-\frac{1}{2}u^2}du\right]
\end{aligned}
$$

We may rewrite this as follows:

$$
\begin{aligned}
f(x) &= e^{\frac{b^2}{2a}}\sqrt{\frac{2\pi}{a}}\left[\frac{1}{\sqrt{2\pi}}\int_{u=\sqrt{a}\left(x-\frac{b}{a}\right)}^{\infty} e^{-\frac{1}{2}u^2}du\right] \\
&= e^{\frac{b^2}{2a}}\sqrt{\frac{2\pi}{a}}\left[1 - N\left(\sqrt{a}\left(x - \frac{b}{a}\right)\right)\right] \\
&= e^{\frac{b^2}{2a}}\sqrt{\frac{2\pi}{a}}N\left(-\sqrt{a}\left(x - \frac{b}{a}\right)\right),
\end{aligned}
$$

where I used the property $[1 - N(z)] = N(-z)$. The skills required to answer this question are very similar to the skills required to manipulate pdfs when deriving option pricing formulae in a Black-Scholes world.

Answer 1.61: Well, if the answer were zero they wouldn't be asking it. Just multiply by the ratio of its conjugate to itself and divide numerator and denominator by x:

$$\begin{aligned}
\lim_{x\to\infty}\left(\sqrt{x^2+x}-x\right) &= \lim_{x\to\infty}\left[\left(\sqrt{x^2+x}-x\right)\cdot\left(\frac{\sqrt{x^2+x}+x}{\sqrt{x^2+x}+x}\right)\right] \\
&= \lim_{x\to\infty}\left[\frac{x^2+x-x^2}{\sqrt{x^2+x}+x}\right] \\
&= \lim_{x\to\infty}\left[\frac{x}{\sqrt{x^2+x}+x}\right] \\
&= \lim_{x\to\infty}\left[\frac{1}{\sqrt{1+\frac{1}{x}}+1}\right] \\
&= \frac{1}{2}
\end{aligned}$$

Answer 1.62: Let $y = x^x$, then take logs, differentiate implicitly, use the product rule, and then use the definition of y to recover the answer:

$$\begin{aligned}
\ln(y) &= x\ln(x) \\
\frac{1}{y}\frac{dy}{dx} &= \ln(x)+1 \\
\frac{dy}{dx} &= x^x(1+\ln(x))
\end{aligned}$$

Answer 1.63: No, of course not. Replace the word "prime" with any other word, and the answer is still no. If they are consecutive, then by definition there are none of them in between!

Appendix B

Derivatives Answers

This appendix contains answers to the questions posed in chapter 2.

Answer 2.1: Most people incorrectly deduce that the call option is worthless. If this is your conclusion, stop looking at the answers and go back and think again. You missed the point. Many people think that zero volatility means the stock price is going nowhere. However, volatility of returns is, by definition, the average deviation from expected returns. It follows that zero volatility means the stock price drifts up at the expected return on the stock with no deviations from this path.

With no volatility, the stock is riskless. In the absence of arbitrage opportunities, the stock must offer an expected return equal to the riskless rate. This is true in both the real world and the theoretical risk-neutral world. This result (expected return equals r) is very strange in the real world—stocks normally offer higher returns. Do note, however, that *all* stocks in the risk-neutral world have expected return equal to the riskless rate. Although I discuss option pricing in the risk-neutral world, the same arguments apply in the real world in the no-volatility case.[1]

The required rate of return on the stock is the riskless interest rate. It follows that with no volatility, the stock price rises to about \$105 for sure.[2] That is, the option finishes in-the-money for sure and is thus riskless. The discounted expected payoff is thus roughly $\frac{(\$105-\$100)}{1.05} = \frac{5}{1.05}$. At 5%, you lose about five cents on every dollar when you discount over a year. The discounted expected payoff is, therefore, about \$4.75, and this is the call value.

This is a good place to mention an often overlooked connection between options and forwards. Suppose that $S(t)$ is the price today of a stock that pays no

[1] If option pricing is done using real world probabilities rather than risk-neutral ones, then the discount rate on the option is a path-dependent random variable that changes as the stock price changes (Arnold and Crack [2004]; Arnold, Crack and Schwartz [2009, 2010]). Such a model allows inference of real world probabilities of a real option project being successful, a financial option finishing in-the-money, or a corporate bond defaulting.

[2] If the interest rate is an effective (i.e., simple) rate, then this is exact; if it is continuously compounded, this is an approximation.

dividends. Let r denote the continuously compounded interest rate per annum. Then a fair price for delivery of the stock at time T is: $F = S(t)e^{r(T-t)}$. In the absence of volatility, the expected time-T stock price is just the forward price. Once volatility is introduced into the picture, the distribution of terminal stock price $S(T)$ becomes spread out. However, the mean of the (risk-neutral) distribution of $S(T)$ is unchanged, and this mean equals the forward price, which is also unchanged: $S(t)e^{r(T-t)}$. That is, the expected time-T stock price in the risk-neutral world is just the forward price.

Back to our option: With no volatility, the value of the option at time t is just the discounted expected time-T payoff in a risk-neutral world:

$$\begin{aligned} c(t) &= e^{-r(T-t)}\max(S(T)-X,0) \\ &= e^{-r(T-t)}\max(S(t)e^{r(T-t)}-X,0) \\ &= e^{-r(T-t)}\max(F-X,0), \end{aligned}$$

where $F = S(t)e^{r(T-t)}$ is the forward price for the stock, and X is the strike price. If follows that the option has value if and only if the forward price exceeds the option's strike.

How do you hedge this? If $F > X$, the option will finish in-the-money for sure, so you need a delta of $+1$. If $F \leq X$, the option will die worthless for sure, so you need a delta of 0 (who would buy the option in this case anyway?).

Answer 2.2: The gamma of an option is the rate of change of its delta, Δ, with respect to stock price—denoted Γ. Option gamma is also called "curvature," or "convexity." Gamma is non-negative for standard puts and calls (their deltas rise with increasing S). Put-call parity tells us that the gamma of a European call is the same as the gamma of a European put.

Option value "decays" toward kinked final payoff as expiration approaches (see figure B.1—first panel). This time decay is called "theta." We usually think about theta as being negative for plain vanilla options, but there are two clear exceptions. A deep in-the-money European-style call can have positive theta if the dividend yield is high enough—because high dividends can push price down below intrinsic value and the option then has to "decay upward" in value as expiration approaches. Similarly, a deep in-the-money European put decays upward in value—because life does not get much better than a deep-in-the-money American put, but the European put cannot be exercised immediately and hence the discount. Crack (2009) discusses this in more detail.

Theta is large and negative for at-the-money options, and it increases in magnitude as maturity approaches. Theta and gamma are typically of opposing signs (the postive theta cases mentioned above are exceptions), so large negative theta typically goes hand-in-hand with large positive gamma. That is, shortening maturity accelerates at-the-money option prices towards the kink and also gives more curvature (i.e., gamma) in the plot of option value as a function of stock price (see figure B.1—third panel).

The maturity/gamma relationship is reversed away from the strike price. If a call is deep in-the-money, then $\Delta \rightarrow 1$, as expiration approaches (for a deep in-the-money *put*, $\Delta \rightarrow -1$ as expiration approaches). Thus, short maturity calls or puts that are deep in-the-money have deltas that do not vary much as S changes. With little variation in delta, the gamma is close to zero. If an option is instead deep *out*-of-the-money, then its gamma is also close to zero because its delta is close to zero with little variation across S. It follows that for away-from-the-money standard options, shorter maturity implies lower gamma for both puts and calls (see figure B.1—third panel).

The gamma (i.e., convexity) for a standard European call on a stock that pays a continuous dividend at rate ρ is given as follows:

$$\Gamma(t) \equiv \frac{\partial^2 c(t)}{\partial S(t)^2} = \frac{e^{-\rho(T-t)-\frac{1}{2}d_1^2}}{S(t)\sigma\sqrt{2\pi(T-t)}},$$

where

$$d_1 = \frac{\ln\left(\frac{S(t)}{X}\right) + (r-\rho+\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

With $(T-t) > 0$, the formula for Γ shows that as $S(t) \rightarrow \infty$, the numerator goes to zero (because $d_1 \rightarrow \infty$), and the denominator goes to infinity. Both limits have the same effect on Γ, pushing it to zero. Similarly, if $(T-t) > 0$, then as $S(t) \rightarrow 0$, $d_1^2 \rightarrow \infty$ so the numerator goes to zero again. However, having S in the denominator pushes Γ in the opposite direction as $S \rightarrow 0$. The exponentiation of d_1^2 in the numerator is much more powerful than the linearity of S in the denominator, so the ratio, Γ, is forced to zero as $S \rightarrow 0$.

If the option is exactly at-the-money [i.e., $S(t) = X$], then as maturity approaches, you have a knife-edge singularity. You get $d_1 \rightarrow 0$, so the numerator of Γ goes to $\mathbf{1}$. However, the denominator tends to zero, so the ratio, Γ, blows up. That is, you get "infinite gamma" at the kink as maturity approaches.

Infinite gamma means the sensitivity of delta to small changes in price of the underlying is infinite. This means that the delta can jump from one-half up to one, or down to zero with just a hair's breadth move in the stock price. In this knife-edge scenario, any delta-hedge that you establish is extremely sensitive to a move in the underlying—you are not hedged.

If you try gamma-hedging (adding traded options to your delta-hedge to replicate the convexity of the derivative), you will need many traded options in your hedge portfolio, and it may become difficult to manage the position.[3] Your problems are similar (but much worse) if hedging barrier options (i.e., "knock-outs") as the price of the underlying approaches the knock-out barrier. The problem is worse near a knock-out's barrier than near a standard call's kink.

[3]In practice, even with a day left to maturity, although the gamma can be quite large, you might need only 10 three-month calls to replicate the convexity of a standard call with one day to maturity—we are not talking infinity here.

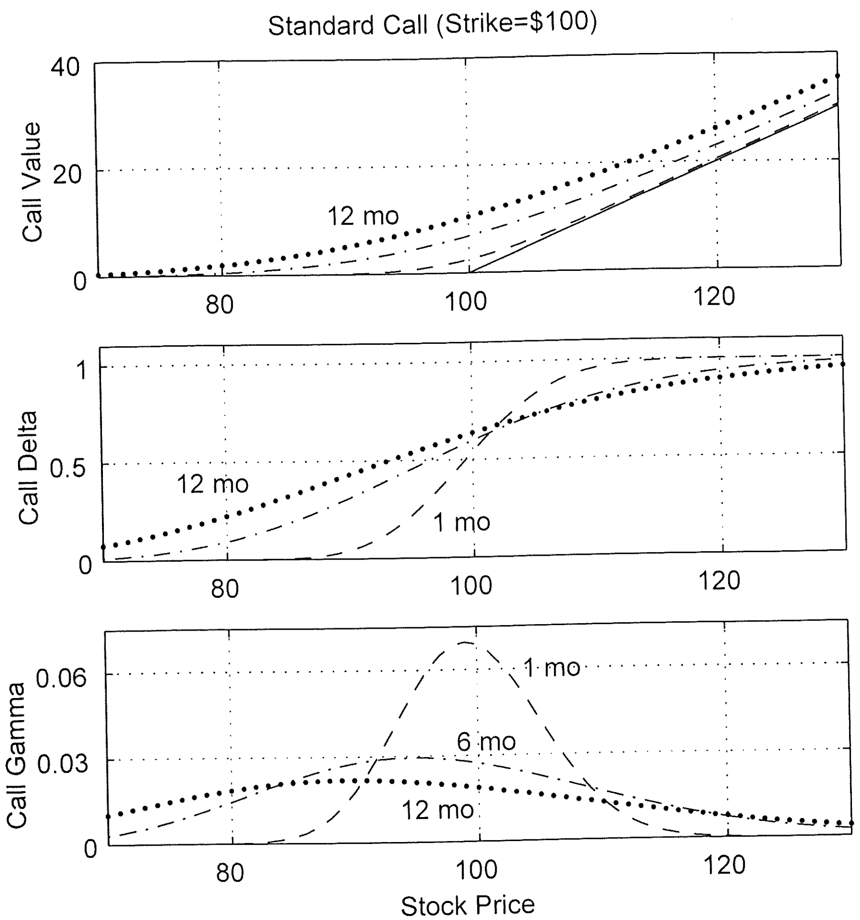


Figure B.1: Standard Call: Price, Delta, and Gamma.

Note: For maturities of 12 months "....", six months "—·—", and one month "– – – –", the call price, delta, and gamma are plotted as a function of price of underlying (see Answer 2.2).

This is because the knock-out's delta can jump from one to zero whereas the standard call's delta jumps only from one-half to zero, or one-half to one.

For American-style options (or more complicated Europeans), you have no closed form formulae. You will probably have to calculate Γ using numerical techniques.

Answer 2.3: The key here is the shape of the risk-neutral distribution of final stock price, $S(T)$, conditional on current stock price, $S(t)$. Many people mistakenly assume the distribution of final stock price to be both symmetric and Normal. The distribution is in fact Lognormal.

The Lognormal distribution is "right skewed," which is also known as "positively skewed." It looks as though its top has been shoved from the right while keeping its base fixed. So, it has a right tail.

If we start with $S(t) = X$, and $r = 0$, then the skewness in the distribution of $S(T)$ means that the final stock price is more likely to end up below the strike than above it.[4] The call has bigger potential payoffs than the put but (because of skewness) lower probabilities of achieving them. The put has smaller potential payoffs than the call but (because of skewness) higher probabilities of achieving them. The bigger payoffs and lower probabilities for the call exactly match the smaller payoffs and higher probabilities for the put. It follows that the put and call have the same risk-neutral expected payoff and, therefore, have the same value. It is straightforward to confirm this equality of values using put-call parity.

Answer 2.4: This is a common question. Stock price, $S(t)$, ranges from \$0 to \$$\infty$; the "delta" varies from 0 to +1. When $S(t)$ is very low (well out-of-the-money), delta is close to zero; when $S(t)$ is very high (well in-the-money), delta is close to one; when $S(t) = X$ (at-the-money), delta is very slightly higher than one-half (assuming no dividends). The curve is smooth and looks very much like a cdf (cumulative distribution function). This is not surprising, given that delta $= N(d_1) = N(d_1(S))$, and $N(\cdot)$ is a cdf, and $d_1(S)$ is an increasing function of S. The delta is illustrated in the second panel in figure B.1.

How about the intuition? The delta is how many units of stock you need to hold to hedge a short call option. If your call option is deep in-the-money, you need one unit of stock because the option will be exercised and the stock will be called; if your option is deep out-of-the-money, you need no stock because the option will expire worthless and the stock will not be called; if your option is at-the-money, you are not too sure, and you have about one-half a unit of stock just in case.

[4] With $r = 0$, the median of the risk-neutral distribution of $S(T)$ conditional on $S(t)$ is $S(t)\ e^{(r-\frac{1}{2}\sigma^2)(T-t)} = S(t)\ e^{-\frac{1}{2}\sigma^2(T-t)} < S(t)$. The option is struck at-the-money [i.e., $S(t) = X$], so the median is below the strike.

Answer 2.5: Without dividends, the standard Black and Scholes (1973) pricing formula for the European call option is given by

$$\begin{aligned} c(t) &= S(t)N(d_1) - e^{-r(T-t)}XN(d_2), \quad \text{where} \\ d_1 &= \frac{\ln\left(\frac{S(t)}{X}\right) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}, \quad \text{and} \\ d_2 &= d_1 - \sigma\sqrt{T-t}. \end{aligned}$$

The option's "delta" is given by $\frac{\partial c(t)}{\partial S(t)} = N(d_1)$. With the option struck at-the-money, $S(t) = X$, and thus, $\ln\left(\frac{S(t)}{X}\right) = 0$ [remember that $\ln(1) = 0$]. All other terms in d_1 are positive. Therefore, $d_1 > 0$, and $N(d_1) > 0.5$ (remember that $N(0) = 0.5$ and $N(\cdot)$ is an increasing function of its argument). Thus, an at-the-money option on a non-dividend-paying stock always has a delta slightly greater than one-half.

Answer 2.6: With continuous dividends at rate ρ, the standard Black-Scholes pricing formula for the European call option is given by[5]

$$\begin{aligned} c(t) &= S(t)e^{-\rho(T-t)}N(d_1) - e^{-r(T-t)}XN(d_2), \quad \text{where} \\ d_1 &= \frac{\ln\left(\frac{S(t)}{X}\right) + (r - \rho + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}, \quad \text{and} \\ d_2 &= d_1 - \sigma\sqrt{T-t}. \end{aligned}$$

The option's "delta" is given by $\frac{\partial c(t)}{\partial S(t)} = e^{-\rho(T-t)}N(d_1)$. With the option struck at-the-money, $S(t) = X$, and thus, $\ln\left(\frac{S(t)}{X}\right) = 0$ [remember that $\ln(1) = 0$]. This, combined with $r > \rho$ yields $d_1 > 0$, and thus $N(d_1) > 0.5$. The naive answer is that $N(d_1) > 0.5$ and that this is the delta—forgetting that $e^{-\rho(T-t)}$ pre-multiplies $N(d_1)$ in the continuous-dividend case. In general, you cannot tell whether the delta, $e^{-\rho(T-t)}N(d_1)$, is larger or smaller than 0.5: it depends upon the size of σ^2. However, in this particular case, $\rho = 0.03$ is so small that $\Delta > 0.5$ for any σ.

Answer 2.7: Almost every person I have asked has got the answer to this one backwards at first. This is unfortunate, because it is a commonly asked question. Think it through carefully before answering, and do not get caught out. The delta is the number of units of stock in the replicating portfolio. Other things being equal, the delta falls with a fall in stock price. However, you are long the call and short the replicating portfolio. This means that the number of units of stock you are short has to fall. So, you must borrow more money and buy back some stock.

[5] This extension of Black-Scholes is due originally to Merton (1973, Footnote 62). Note, however, that his original formula has an obvious typo in it (he omits the dependence of d_1 and d_2 on ρ).

If you got it wrong, think about it as follows. Ask yourself how the replicating portfolio changes (e.g., delta falls, so less stock is needed in the replicating portfolio). Then ask yourself whether you are long or short the replicating portfolio (you are short here). If you are short, be sure to reverse the implications (less stock shorted means you must borrow to buy some back).

Answer 2.8: With the standard European call, you have a simple closed-form expression for the option's delta. For example, (under the Black and Scholes [1973] assumptions) the delta of a standard European call on a non-dividend-paying asset is equal to $N(d_1)$ where

$$d_1 = \frac{\ln\left(\frac{S(t)}{X}\right) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}.$$

See Answer 2.6 for the delta in the case where there are continuous dividends at rate ρ.

Unfortunately, only a few known options have closed-form pricing formulae. For exotic options with no closed-form pricing formula, you need a pricing algorithm. This may be a Monte-Carlo simulation,[6] a binomial tree, a numerical PDE solution routine, or perhaps an ODE approximation to a PDE. By varying the input value of the current level of underlying, you can use the pricing algorithm to calculate a numerical derivative of price with respect to level of underlying; i.e., the delta. All you are doing is using the computer rather than the calculus to tell you how the option price changes with a change in stock price.

Answer 2.9: The pricing formula for the standard Black-Scholes European call option on a non-dividend-paying stock is:

$$c(t) = S(t)N(d_1) - e^{-r(T-t)}XN(d_2),$$

where d_1 and d_2 are as previously defined. $N(d_1)$ is the option's "delta," sometimes denoted "Δ." $\Delta = N(d_1)$ is the same thing as the partial derivative of call price with respect to underlying: $\frac{\partial c(t)}{\partial S(t)}$. It measures how the call price changes per unit change in the price of the underlying.

[6] As an introduction to exotic options and Monte-Carlo techniques, I recommend the Monte-Carlo chapter of my book *Basic Black-Scholes* (Crack [2009]). The earliest Monte-Carlo reference I know of in option pricing is Boyle (1977). Boyle also gives techniques for accelerating the convergence of Monte-Carlo estimation and some references to the mathematics literature (see Hull [1997, pp365–368] for other techniques). For background information on the development of exotics and the players in the market, see Fraser (1993); for a slightly higher-level than Hunter and Stowe (1992), see Ritchken, Sankarasubramanian, and Vijh (1993) or Hull (1997); at a slightly higher level still, see Goldman, Sosin, and Gatto (1979) or Conze and Viswanathan (1991). Note that the value of a look-back option to buy at the minimum or sell at the maximum might arguably be considered an upper bound on the value of market timing skills—see Goldman, Sosin, and Shepp (1979) for more details.

Another interpretation of the terms involves a replicating portfolio. $\Delta = N(d_1)$ is the number of units of stock you must hold in a continuously rebalanced portfolio that replicates the payoff to the call. The term $e^{-r(T-t)}XN(d_2)$ is the value of the borrowing (or a short position in bonds) required in a continuously rebalanced portfolio that replicates the payoff to the call. The value of the borrowing in the replicating portfolio is always less than or equal to the value of the replicating portfolio's long position in the stock. This is equivalent to stating that the call has non-negative value.

Another interpretation of the terms involves expected benefits and expected costs to owning the call. The term $S(t)N(d_1)$ is the discounted value of the expected *benefit* of owning the option (expectations taken under a risk-neutral probability measure). Why is the $N(d_1)$ there? Well, $N(\cdot)$ is a cumulative density function, so it must be that $N(d_1) \leq 1$. This in turn implies that $S(t)N(d_1) \leq S(t)$. This is because the future benefit of owning the option is $S(T)$ if the option finishes in-the-money and zero if it finishes out-of-the-money (or "under water"). This benefit is strictly dominated by a long position in the stock (a position that returns $S(T)$ regardless of whether the option is in- or out-of-the-money and costs $S(t)$ now). It follows that you value the benefit from the call at less than the long position in the stock, $S(t)N(d_1) \leq S(t)$. It is for this reason that the $N(d_1)$ term multiplies the $S(t)$ term.[7]

The term $e^{-r(T-t)}XN(d_2)$ is the discounted value of the expected *cost* of owning the option (with expectations taken under a risk-neutral probability measure). You can see all the components of the discounted expected value as follows: $N(d_2)$ is the (risk-neutral) probability that the call option finishes in-the-money (see extended discussion in Crack [2009]); X is your cost if it does; and $e^{-r(T-t)}$ is the discounting factor.

Here is a summary of the foregoing paragraphs (where "$P(in)$" denotes the risk-neutral probability that the call finishes in-the-money):

$$c(t) = \overbrace{S(t)\underbrace{N(d_1)}_{\Delta}}^{\text{stock position \& benefit}} \quad \underbrace{\overbrace{-\,e^{-r(T-t)}X\underbrace{N(d_2)}_{P(in)}}^{\text{bond position}}}_{\text{borrowing \& cost}}$$

The value of the standard European put on a non-dividend-paying stock may now be deduced. The present value of the *benefit* of owning the put is $e^{-r(T-t)}X[1-N(d_2)]$, where $[1-N(d_2)]$ is the (risk-neutral) probability that the put option finishes in-the-money (i.e., the call finishes out-of-the-money), X is your payoff if it does, and $e^{-r(T-t)}$ is the discounting factor.

[7] The first term is $S(t)N(d_1) = e^{-r(T-t)}E^*[S(T)\mathcal{I}_{S(T)>X}|S(t)]$, where E^* denotes expectation taken with respect to the risk-neutral probability measure, and $\mathcal{I}_{S(T)>X}$ is as given in equation B.1.

$$\mathcal{I}_{S(T)>X} = \begin{cases} 1 \text{ if } S(T) > X, \\ 0 \text{ if } S(T) \leq X. \end{cases} \tag{B.1}$$

The present value of the *cost* of owning the put option is $S(t)[1-N(d_1)]$. There are two probabilistic interpretations of $N(d_1)$, each under a competing martingale measure (see Crack [2009]).

Using the property that $[1-N(z)] = N(-z)$, the value of the put option must be

$$p(t) = e^{-r(T-t)}XN(-d_2) - S(t)N(-d_1),$$

where d_1 and d_2 are as already defined for the call.

Put-call parity says that

$$S(t) + p(t) = c(t) + Xe^{-r(T-t)} + D.$$

If you plug in $c(t) = S(t)N(d_1) - e^{-r(T-t)}XN(d_2)$, and $D = 0$, you do indeed get that $p(t) = e^{-r(T-t)}XN(-d_2) - S(t)N(-d_1)$, as deduced above.

See Crack (2009, chapter 8) for extensive discussion of Black-Scholes interpretations and intuition.[8]

Answer 2.10: Questions about a "digital option" or "binary option" are quite common. The digital "cash-or-nothing" option that pays H if $S(T) > X$ has a value of $He^{-r(T-t)}N(d_2)$. This is simply the discounted (risk-neutral) expected payoff to the option: $N(d_2)$ is the (risk-neutral) probability that the option finishes in-the-money; H is the payoff if it does; and $e^{-r(T-t)}$ is the discounting factor. H is sometimes called the "bet." If H is chosen to equal the strike of the standard Black-Scholes option, then the cash-or-nothing option has the same value as the second term in the Black-Scholes formula: $e^{-r(T-t)}XN(d_2)$.

The first term in the Black-Scholes formula, $S(t)N(d_1)$, is the value of a long position in a digital "asset-or-nothing" option. A long position in the asset-or-nothing option, combined with a short position in the cash-or-nothing option, replicates the payoff to the European call—and, therefore, has the same value (you should draw the payoff diagrams to verify this).[9]

Be sure to see Question 2.11 and Answer 2.11 for more details on the binary option.

Answer 2.11: I look at this intuitively first and then more rigorously. Intuitively, if the digital "cash-or-nothing" option is deep in-the-money, you are just waiting for your fixed cash payoff, and increases in volatility can only decrease your payoff. If you are deep out-of-the-money, you are expecting nothing, and

[8] Note that there is a competing stock-numeraire world where if the bond de-trended by the stock follows a martingale, then $N(d_1)$ in the Black-Scholes formula is the probability that the call finishes in the money (see Crack [2009] for details). It is a Z-score argument similar to the one that establishes $N(d_2)$ as the probability that the call finishes in the money in a world in which the stock de-trended by the bond follows a martingale.

[9] As an aside, you might like to note that the payoff to the European call may also be replicated by using barrier options: you need a "knock-out" call option plus a "knock-in" call option.

increases in volatility can only increase your payoff. If $c(t)$ is the price of the digital cash-or-nothing option, then somewhere around the at-the-money position, the sign of $\frac{\partial c(t)}{\partial \sigma^2}$ must change.

Rigorously, if $c(t)$ is the price of the digital cash-or-nothing option, then direct calculation (under Black-Scholes assumptions) shows that

$$\frac{\partial c(t)}{\partial \sigma^2} > 0 \text{ if and only if } S(t) < Xe^{-(r+\frac{\sigma^2}{2})(T-t)}.$$

Another (equivalent) way of looking at this is that $\frac{\partial c(t)}{\partial \sigma^2} > 0$ if and only if the probability of finishing in-the-money increases with an increase in σ^2, and this is so if and only if $S(t) < Xe^{-(r+\frac{\sigma^2}{2})(T-t)}$.

Figure B.2 (on page 115) shows $\frac{\partial \text{ CALL PRICE}}{\partial \sigma^2}$ for the asset-or-nothing digital option, the cash-or-nothing digital option, and the standard call (all options are European). The price of the standard call is just the difference between the prices of the asset-or-nothing digital option and the cash-or-nothing digital option. Differentiation is a linear operation, so the sensitivity of the standard call to volatility is just the difference between the sensitivity of the asset-or-nothing digital option and the cash-or-nothing digital option.

It is clear from figure B.2 that the price of the standard call is increasing in volatility. This should come as no surprise. A call option is an insurance policy. It puts a floor on your losses. When there is more risk about, the premium (i.e., call price) should be higher. In the same way, you should be happy to pay more for fire insurance if you find out that your next-door neighbor is an arsonist. See Chance (1994) for more details on the sensitivity of option value to the various input parameters.

For the cash-or-nothing, the boundary on the sign of $\frac{\partial c(t)}{\partial \sigma^2}$ is always slightly less than X (see figure B.2 for a clear illustration). Thus, if you are in-the-money, or at-the-money, more volatility is always bad; if you are very slightly out-of-the-money, more volatility is still bad [when $Xe^{-(r+\frac{\sigma^2}{2})(T-t)} < S(t) \leq X$]; if you are well out-of-the-money, more volatility is always good [when $S(t) < Xe^{-(r+\frac{\sigma^2}{2})(T-t)}$]. This differs from the standard European call option for which $\frac{\partial c(t)}{\partial \sigma^2}$ is always non-negative (see figure B.2).

You might ask why the boundary on the sign of $\frac{\partial c(t)}{\partial \sigma^2}$ is always slightly less than X, rather than exactly at X. The relationship between volatility and skewness in the (Lognormal) distribution of final stock price is where the explanation lies. There are two forces at work: First, an increase in σ^2 tends to spread out the distribution of $S(T)$, putting more probability weight into the tails; second, increasing σ^2 drags down the median of the distribution,[10] tending to pull probability weight leftward and out of the right tail, thus increasing

[10] The mean of the risk-neutral distribution of $S(T)$ conditional on $S(t)$ is $S(t)e^{r(T-t)}$; the median is $S(t)e^{(r-\frac{1}{2}\sigma^2)(T-t)}$.

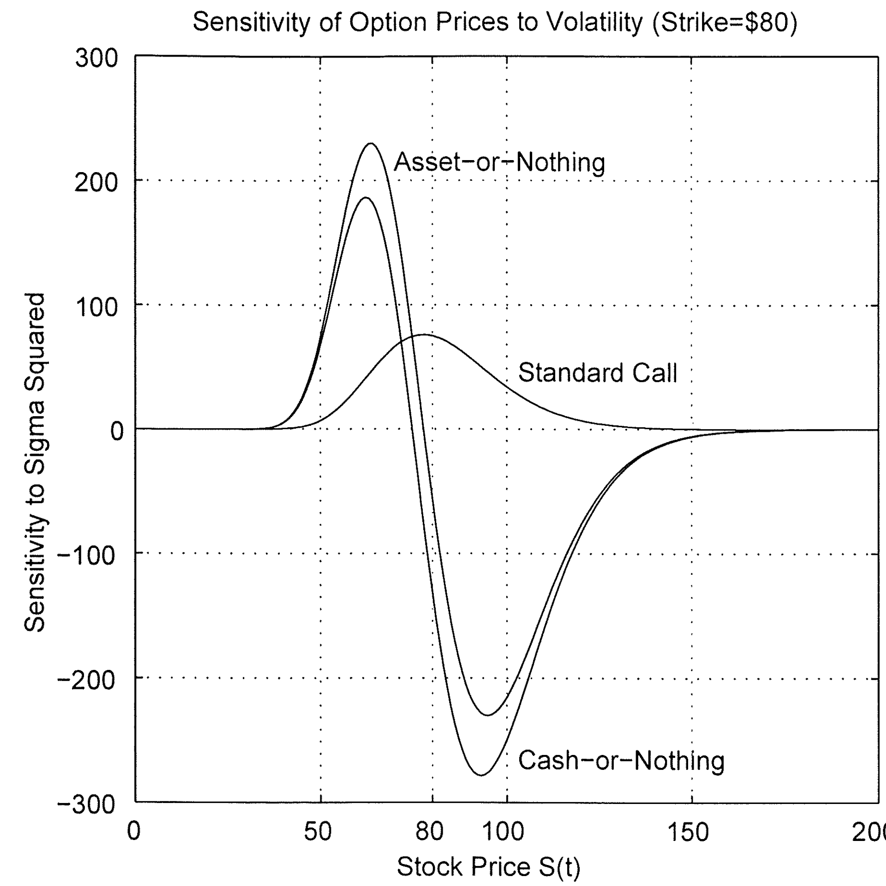


Figure B.2: Sensitivity of Option Prices to Volatility

Note: The figure plots $\frac{\partial \text{ CALL PRICE}}{\partial \sigma^2}$ (i.e., "vega") using parameters $X = 80$, $r = 0.05$, $T - t = 1$, and $\sigma = 0.20$. The asset-or-nothing call price is always more sensitive to σ^2 than the cash-or-nothing call price. The difference between the sensitivities of each digital option is thus non-negative. The standard European call is equivalent to a long position in the asset-or-nothing and a short position in the cash-or-nothing. The response of the standard call price to increases in σ^2 is thus non-negative.

the skewness. If the strike price is at or below the median of $S(T)$ (so the option is in-the-money, or very slightly out-of-the-money), then both forces push probability mass leftward, increasing the likelihood of finishing out-of-the-money. However, if the strike price is far above the median of $S(T)$ (so the option is far out-of-the-money), then the increasing spread of the distribution dominates the leftward move of the median, and the probability of finishing in-the-money increases with increasing σ^2. For the forces to be balanced, the option must be struck above the median of the distribution of $S(T)$. The strike price that just balances the influence of both forces is $X = S(t)e^{(r+\frac{\sigma^2}{2})(T-t)}$. At this strike, the option is insensitive to instantaneous changes in σ^2 and it is slightly out-of-the-money: $S(t) = Xe^{-(r+\frac{\sigma^2}{2})(T-t)}$.

Answer 2.12: The naive and time-consuming way to find the delta for the knock-out option (or "barrier option") is to differentiate the closed-form pricing formula for the down-and-out, find $\frac{\partial C}{\partial S}$, and compare it to the same quantity for the standard call.[11] It is more elegant to use common sense and some limiting relationships to deduce the relationships between the deltas of the knock-outs and the standard call.

The delta is the sensitivity of call price to underlying. This means that the option's delta is just the slope when you plot call value, $c(t)$, against underlying value, $S(t)$. Do not get this plot mixed up with the *payoff* diagram (the one with a "kink" for a standard call). See figure B.1 (on page 108).

Now, everything you can do with a down-and-out option, you can also do with a standard option. On top of that, you still have a standard option in your hands in cases where the down-and-out gets "knocked out." It follows that the standard call is more versatile than the down-and-out call and must be more expensive. Thus, the value of the standard call must plot above the value of the down-and-out call for any value of the underlying. However, the two calls have the same value for very large $S(t)$—because the down-and-out option is unlikely to get knocked out. Both valuation curves are smooth, so the down-and-out call's valuation curve must be steeper [it starts lower than the standard call and "finishes" in the same place for high $S(t)$]. A steeper valuation curve when plotted against the level of underlying means precisely that the delta is higher for the down-and-out call than for the standard call.

For the up-and-out, you get a different answer. As before, the up-and-out is a knock-out option and is cheaper than the standard call. However, the standard call option and the up-and-out call option have the same value for very small $S(t)$—because the up-and-out option is unlikely to get knocked out. Both valuation curves are smooth, so the up-and-out call's valuation

[11]The closed-form valuation formula for the down-and-out, together with discussion, is in Merton ([1973, pp175–76]; [1992, p302]). It takes around 15 minutes to differentiate it by hand carefully and about the same time to program the numerical derivative in MATLAB. The down-and-out option was introduced by Gerard Snyder (1969). See his paper for a look at the operations of the options markets in the late sixties.

curve must be flatter (it starts in the same place as the standard call and finishes lower). A flatter valuation curve means precisely that the delta is lower for the up-and-out call than for the standard call.

Thus, the following relationships hold for the deltas of the different options:

$$\Delta_{up-and-out\ call} \leq \Delta_{standard\ call} \leq \Delta_{down-and-out\ call}$$

To hedge a short position in a down-and-out call, you need to buy more units of stock than you do to hedge a short position in a standard call. The value of the down-and-out call is more sensitive than the standard call to changes in the value of the underlying stock.

Note that increasing the term to maturity or increasing the knock-out price both increase the likelihood that a down-and-out call will be knocked out. This makes the down-and-out call even cheaper relative to the standard call. In fact, if the down-and-out call is very likely to be knocked out, the plot of down-and-out call price against stock price can become concave. Conversely, if the term to maturity is very short and the knock-out price is very low, the standard call and the down-and-out call have virtually identical prices (because the knock-out is very unlikely to be knocked out).

Answer 2.13: Your observation is that the sample variances are not linear in time and that the differences are statistically significant. This is equivalent to rejecting the null hypothesis of a random walk using a "variance ratio" test (Lo and MacKinlay [1988]).[12] This is contrary to the random walk assumptions of the Black-Scholes model.

The observations are consistent with the empirical findings that some financial stock indices are positively autocorrelated at weekly return intervals (Lo and MacKinlay [1988]).[13] This predictability influences the theoretical value and the empirical estimate of the diffusion coefficient σ (Lo and Wang [1995]). An adjustment can be made to the Black-Scholes formula to account for the predictability that is not part of the original Black-Scholes model. A new diffusion process that captures the predictability can be defined (Lo and Wang [1995]).

With the new specification, the autocorrelation is described using a more complicated drift in the diffusion. The drift is now important for pricing the option. In the old specification, drift was not important (Black and Scholes [1973]; Merton [1973]).

The final pricing formula takes the same form as the original Black and Scholes (1973) formula. However, the way in which the volatility term σ is estimated

[12]See also Peterson et al. (1992) for related variance ratio testing in the commodities market; their findings lead them to a brief discussion of option pricing in the presence of autocorrelation.

[13]Autocorrelation in a time series is correlation between observations and themselves lagged. It is also known as "serial correlation." Its presence neither implies, nor is implied by, the presence of a drift. Consult your favorite statistics book for more information.

changes. An increase in autocorrelation may either increase or decrease the value of σ—it depends upon the specification of the drift (Lo and Wang [1995, p105]).

The presence of autocorrelation in stock returns is only one example of a real world divergence from the Black and Scholes (1973) assumptions. For example, Thorp (1973) discusses the effect of restrictions on short sales proceeds. See Hammer (1989) for a discussion of other deviations.

Answer 2.14: With no dividends, it is never optimal to exercise the plain vanilla American call option prior to maturity because the option is worth more "alive" than "dead." If you never exercise early, then the "American" feature of the call is not valuable. Thus, the standard American call option and European call option (on a non-dividend-paying stock) have equal values. See Crack (2009) for extensive discussion.

Figure B.3 plots the time value of the call option, $c(t) - \max[S(t) - X, 0]$, against $S(t)$ for the parameter values $X = 80$, $r = 0.05$, $T - t = 1$, and $\sigma = 0.20$. I have replaced the American call value $C(t)$ with the European value $c(t)$ because they are the same thing for plain vanilla call options in the absence of dividends.

The time value (the height in figure B.3) tends to zero as expiration approaches, and this is regardless of stock price. The existence of positive time value (i.e., value over and above exercise value) means that there is value in waiting to exercise. It is this value that makes the American call more valuable alive than dead. However, this does not mean that you should continue to hold the option. Rather, it means that if you wish to exit the call position, you should sell it, not exercise it. The time value is easily seen by looking at the excess of option price over intrinsic value in the "Listed Options Quotations" (i.e., options on individual stocks) in the *Wall Street Journal*.

How does time value arise? There are costs to exercising a call prior to maturity: you lose the interest you would have earned on the strike price and you lose the ability to exercise later. These costs are both intimately linked with the time to maturity, and thus they decline to zero as maturity approaches. There is a benefit to early exercise of a call: you capture any dividend payment on the underlying. In the presence of dividends, you gain the benefits with least cost by waiting until just prior to the ex-dividend day to exercise. In this case, you would exercise only if the benefit outweighs the costs. In practice, these costs of early exercise typically outweigh the benefit until the last ex-dividend date during the life of the option (Cox and Rubinstein [1985, p144]). By this time the costs of early exercise have depreciated substantially. A very large expected dividend might also trigger early exercise.

Answer 2.15: The naive answer is that as stock price falls, so too does the delta. However, this ignores the influence of the passage of time on your hedge. This is a good question, because you must think in both dimensions.

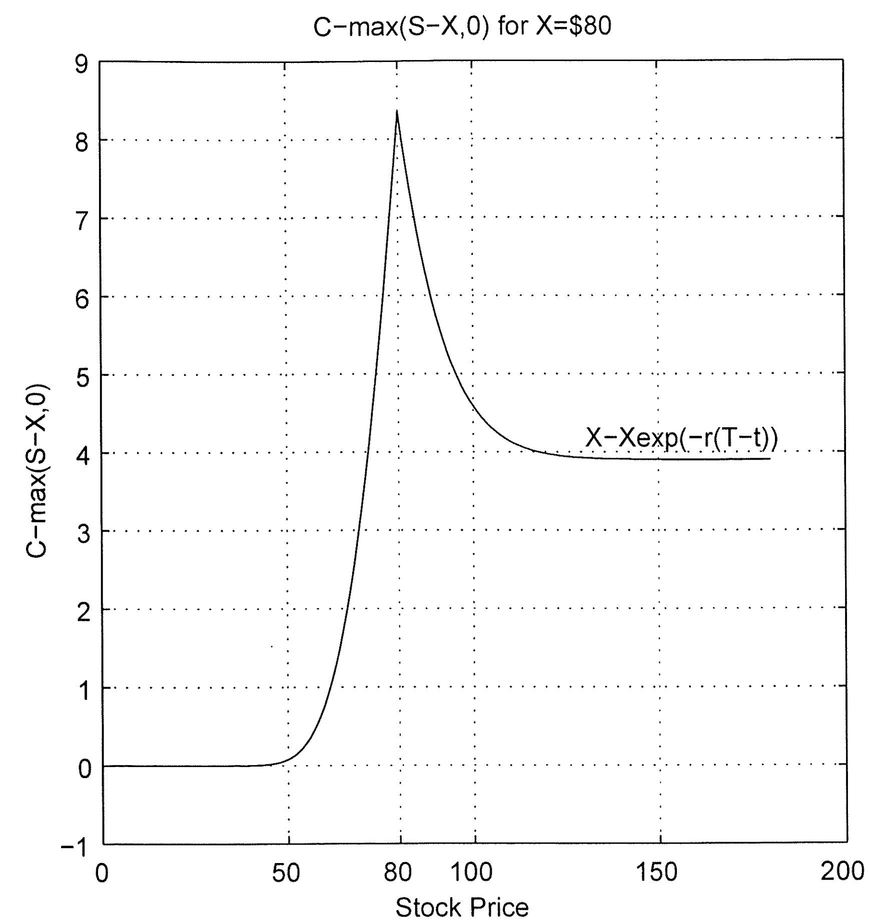


Figure B.3: Time Value of a European Call Option

Note: The difference $c(t) - \max[S(t) - X, 0]$ is the value of not exercising. When the option is deep out-of-the-money, $\max[S(t) - X, 0] = 0$, and $c(t)$ is approximately zero. When the option is deep in-the-money, you save X by not exercising now, but it costs you the present value of exercising at maturity: $X \times e^{-r(T-t)}$. The left-hand limit of $X - X \times e^{-r(T-t)} = X \times \left(1 - e^{-r(T-t)}\right)$ is always non-negative. The "kink" in $\max[S(t) - X, 0]$ puts the "cusp" in the plot of $c(t) - \max[S(t) - X, 0]$ versus $S(t)$ at $S(t) = X$.

Two opposing forces are at work here: First, other things being equal, the delta of a call option that is in-the-money rises toward +1 as the option gets closer to expiration;[14] second, other things being equal, as stock price falls, the delta of a call option falls.

If stock price is observed to fall gently over the final two months, and the option remains in-the-money, the approach of the expiration date pushes the delta up to +1. If the fall in stock price is a little stronger, you may see the delta fall somewhat initially (and you will sell stock in your hedge portfolio). However, if the option finishes in-the-money, then the delta rises to +1 at the end of the life of the option (and you will buy stock in your hedge portfolio).

Answer 2.16: This is a good question. Introductory courses typically do not say much about jump processes.

The Black and Scholes (1973) model naively assumes that stock prices are continuous. That is, they assume that you can draw the price history without lifting your pencil from the paper. You need only stand on the floor of an exchange,[15] watch a real-time feed (e.g., on a Bloomberg terminal), or read the WSJ headlines after an "event" to see that prices do not move smoothly. Indeed, the fact that stock prices are typically quoted with a minimum tick size (either exchange-imposed or effective) means that stock prices *cannot* move continuously. You can think of big stock price jumps as being stock price responses to the arrival of information in the market; small stock price jumps might just be due to the random ebb and flow of non-information-based (i.e., liquidity-related) transactions.

A "jump" price process is a price process that has infrequent jumps (i.e., discontinuities) in it. If the jump process is a very simple one, the Black-Scholes/Merton no-arbitrage technique can still be used to hedge and price options on an asset whose price follows the process. If the jump process is more complicated, the no-arbitrage technique breaks down. See the following discussion, and go to the references if you need more details. I have included some lengthy comments and references. This is because I think it is relevant, and it is often not covered in introductory courses.

[14] If the option you sold finishes in-the-money, you need to be long the stock because it will be called away. Of course, if the option is out-of-the-money, the approach of the expiration pushes the delta down to zero. If the option is at-the-money, then (assuming a non-dividend-paying stock) the delta tends to a number slightly greater than one-half as the expiration date approaches (Cox and Rubinstein [1985, figure 5-13, p223]).

[15] I have been on the floors of the New York Stock Exchange (NYSE), Chicago Board of Trade (CBOT), Boston Stock Exchange (BSE), and Dunedin Stock Exchange—long since replaced by screen trading—during trading hours. I have also visited the Chicago Board Options Exchange (CBOE), the Chicago Stock Exchange (CSE), and the old Paris Bourse. The financial futures floor at the CBOT is big enough to fit a 747 jumbo jet with space to spare—and it is noisy as hell. Conversely, the BSE is small and quieter than your typical MBA computer lab. I forecast that all the Chicago futures exchange floors will soon be deserted—replaced by electronic trading. The NYSE may take a little longer, but I think it will suffer the same fate.

A simple jump process example (that is *not* a diffusion) has $\frac{dS}{S} = \mu dt + (J-1)d\pi$ (Cox and Ross [1976, p147]). In this example, $J-1$ is the jump amplitude (where $J \geq 0$), $d\pi$ takes the value +1 with probability λdt and 0 with probability[16] $1 - \lambda dt$. The percentage stock price change $\frac{dS}{S}$ can thus jump suddenly to $J-1$ (which may itself be random); such a jump pushes S to SJ.

In this simple example, if J is fixed (i.e., non-random), a riskless hedge portfolio *can* be formed, and options on an asset whose price follows this simple jump process *can* be valued using the Black-Scholes/Merton no-arbitrage technique. This should come as no big surprise. The only real difference between this "pure Poisson process" case, and the simple binomial option pricing situation (Sharpe [1978]; Cox, Ross, and Rubinstein [1979]; Rendleman and Bartter [1979]; Cox and Rubinstein [1985]; Crack [2009]) is that the arrival time of the jump up or jump down is a random variable. You do not need to know *when* the stock price will jump to hedge the risk in a binomial setting. This "pure Poisson process" is a special case of a more general jump diffusion process discussed next.

Consider the jump diffusion process $\frac{dS}{S} = (\alpha - \lambda k)dt + \sigma dZ + dq$ (described in detail in my Footnote 8 to Question 2.16 on page 24). When $\sigma = 0$ and $Y \equiv dq+1$ is non-random, you get Cox and Ross's simple jump process above, and the no-arbitrage technique can be used to hedge and price options on the jump process.[17] Otherwise, when $\sigma > 0$ and $\mathrm{var}(Y) \geq 0$ it is not possible to form a riskless hedge portfolio or use the no-arbitrage technique (Cox and Ross [1976, p147]; Cox and Rubinstein [1985, pp361–371]; Merton [1992, p316]). Both the (non-jump) diffusion process and the (non-diffusion) simple jump process are the continuous limits of discrete binomial models. However, the jump-diffusion is not. It is for this reason that a riskless hedge cannot be formed in the jump-diffusion case (Cox and Rubinstein [1985, pp361–371]).

The fundamental reason that the no-arbitrage technique can be used to hedge and price options in the standard Black-Scholes world is linearity. In continuous time, the Black-Scholes option price is an instantaneously linear function of the stock price. Portfolio building is a linear operation, and it follows that payoffs to the option can be perfectly replicated by building and continuously rebalancing a portfolio of the stock and the bond. Linearity breaks down when the jump term has positive variance—the call price becomes a nonlinear function of the stock price and perfect hedging is not possible (Merton [1992, p316]).

Although the no-arbitrage technique fails to price the option on the jump diffusion process, you can price the option using an *equilibrium* argument. An

[16] In this example, π is a continuous time "Poisson process." The term λ is the "intensity" of the process.

[17] I thank John Cox for explaining to me why such jump processes can be perfectly hedged (personal communication [February 17, 1994]).

instantaneous CAPM (capital asset pricing model) approach may be used—as it was in the original Black and Scholes (1973) paper. The information that causes jumps may be assumed to be firm-specific (i.e., unsystematic and diversifiable).[18] You can hedge out the non-jump part of the option and deduce that the remainder (the jump) must have zero beta and, therefore, a riskless rate of return. This yields a partial differential equation that can be solved to give the call option price as an infinite summation:

$$C(S(t), T-t) = \sum_{n=0}^{\infty} \left\{ \frac{\exp[-\lambda(T-t)][\lambda(T-t)]^n}{n!} \times E_n\{W[S(t)X_n \exp(-\lambda k(T-t)), (T-t); X, \sigma^2, r]\} \right\}$$

Here X_n is a random variable with the same distribution as the product of n independent and identically distributed random variables each identically distributed to the random variable Y (recall that $Y-1$ is the random percentage change in stock price when a jump occurs), $X_0 \equiv 1$, E_n is the expectation operator over the distribution of X_n, and $W[S, (T-t); X, \sigma^2, r]$ is the standard Black-Scholes pricing formula (see Merton (1992, pp318–320) for a full discussion of the foregoing and Haug [2007, section 6.9.1] for practical issues).

You cannot perfectly hedge the call when the underlying follows the general jump diffusion [$\sigma > 0$, $\text{var}(Y) \geq 0$]. However, you can hedge out the continuous parts of the stock and option price movements. This leaves a risky hedge portfolio following a pure jump process (with stochastic jump size). If you follow the Black-Scholes hedge when you are short the option, then most of the time you earn more than the expected rate of return on the risky hedge portfolio. However, if one of those occasional jumps occurs (i.e., news arrives), you suffer a reasonably large loss. In the non-diversifiable jump case, the return to the hedge portfolio when there is a jump balances the return during normal time to some extent, but not well enough to make the equilibrium return on the hedge equal to the riskless rate; as mentioned above, the hedge is risky.

In general, there is no way to adjust the parameters of the hedge technique (σ^2, for example) to get a better hedge (see Merton [1992, pp316–317] for a full discussion of the issues).[19]

Finally, if the underlying asset price is modeled as a jump process, the standard Black-Scholes call option formula mis-prices the option. Both the magnitude and the direction of the mis-pricing of the Black-Scholes model relative to the jump model vary with the distributional assumption for the size of the jump component (Trippi et al. [1992]).

[18] Note that in situations where the size of the jump is assumed to be systematic, the risk-neutral pricing technique cannot be used to value options. Hull (1997, p449, Footnote 14) directs the reader to Naik and Lee (1990) for a discussion of this point.

[19] For theoretical and empirical comparisons of the Merton (1976) jump process call option pricing and the standard Black-Scholes pricing, see Ball and Torous (1985).

Answer 2.17: Most people upon whom I have tested this one make several mistakes. Please note that at time t prior to maturity the call price function is *not* asymptotic to the line with slope 1 rising from the strike price;[20] Merton (1973) demonstrates that this is not true. If you made this mistake, stop reading here and go back and try again.

The correct plots appear in figure B.4. The parameters used are $X = 80$, $r = 0.05$, $T - t = 1$, and $\sigma = 0.20$. The plot of call value against terminal stock price is the classic "kinked" call option payoff (the top plot in figure B.4). Call value (terminal payoff) rises with slope 1 from the point $S(T) = X$.

The plot of call price versus futures price is a smooth curve that is asymptotic to the line $C = 0$ when the futures price, $F(t, T)$, is very small, and asymptotic to a line rising with slope $e^{-r(T-t)}$ (i.e., slightly less than 1) from the point $F(t, T) = X$ when the futures price is very large (the middle plot in figure B.4). One way to confirm the slope of this line is to use the chain rule: $\frac{\partial C}{\partial F} = \frac{\partial C}{\partial S} \cdot \frac{\partial S}{\partial F}$, where $\frac{\partial C}{\partial S}$ is the delta (which goes to 1 when F and S are large) and $\frac{\partial S}{\partial F}$ is just $e^{-r(T-t)}$. The slope of the line in this middle plot thus gets closer to one as maturity approaches, but it is strictly less than one at any prior time.

The plot of call price versus stock price, $S(t)$, is a smooth curve that is asymptotic to the line $C = 0$ when the stock price, $S(t)$, is very small, and asymptotic to the line that rises with slope 1 from the point $S(t) = Xe^{-r(T-t)}(= \$76.10$ here) when the futures price is very large.[21] See the bottom plot in figure B.4. The last two results are tied together by the fact that $F(t, T) = X \Leftrightarrow S(t) = Xe^{-r(T-t)}$.

At time t prior to maturity, the call price is lower if the futures price is equal to \$10 than it is if the stock price is equal to \$10. This is because the futures price represents expected future value in some sense, and this is not worth as much as current value (\$10 today is worth more than \$10 tomorrow).

Answer 2.18: This is a fundamental question. If it takes you more than five seconds to answer this, you are in trouble. Black and Scholes (1973) assume an arithmetic Brownian motion in log price. This assumption yields a geometric Brownian motion in price and an arithmetic Brownian motion in continuously compounded returns. Volatility of continuously compounded stock returns, σ^2, grows linearly with time for an arithmetic Brownian motion. The four-year σ^2 is four times the one-year σ^2. It follows that the four-year σ is two times the one-year σ. The answer is, therefore, 30%.

If $r > 0$, you must adjust the value of r (r is four times as large when one period is four years as compared to when one period is one year).

[20] A curve is "asymptotic" to a line (i.e., an asymptote) if the curve gets closer and closer to the line. For example, $y = \frac{1}{x}$, for $x > 0$ is asymptotic to the line $y = 0$ as $x \to \infty$ and asymptotic to the line $x = 0$ as $y \to \infty$.

[21] Note that this implies that the time value $\{c(t) - \max[S(t) - X, 0]\} \to \left\{X - Xe^{-r(T-t)}\right\}$ as $S(t) \to \infty$. See figure B.3 (on page 119) for a plot of time value versus stock price.

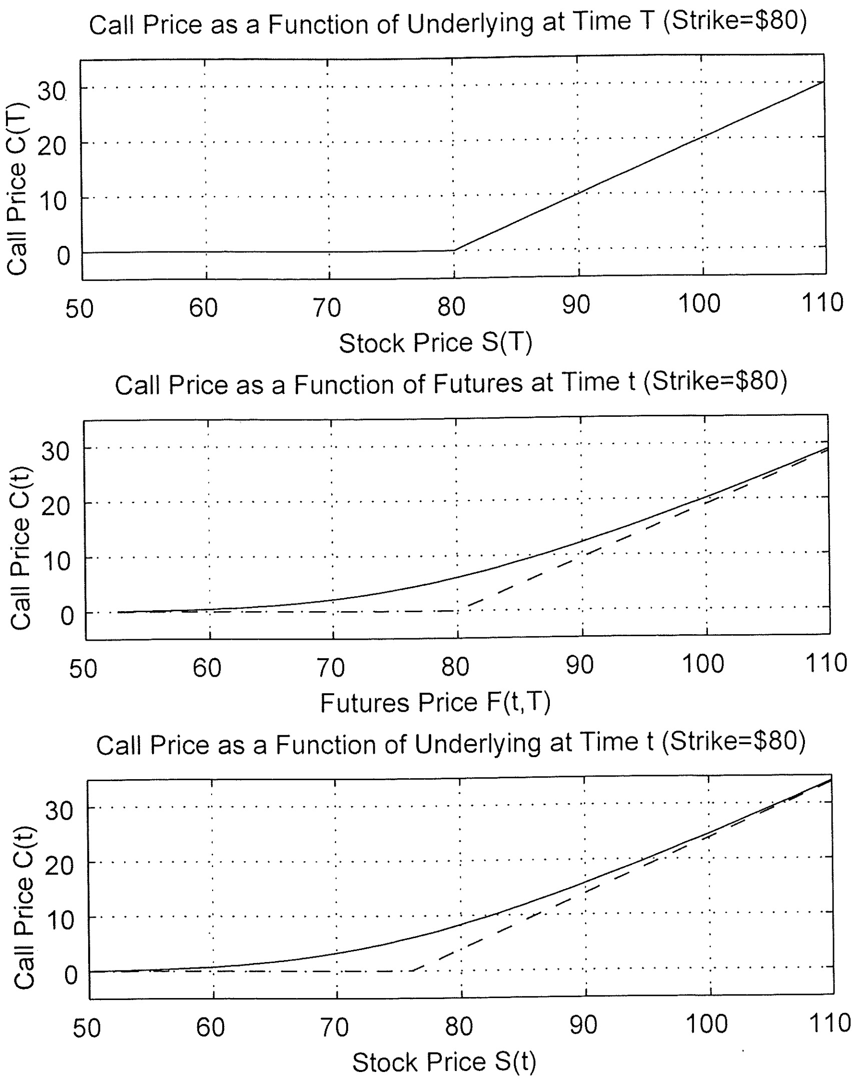


Figure B.4: Call Price as a Function of Different Variables

Note: Call price is plotted as a function of price of underlying and of futures price (see Answer 2.17).

Answer 2.19: Suppose that the process $\mathcal{S}(t)$ is an arithmetic Brownian motion of form

$$d\mathcal{S}(t) = \mu dt + \sigma_A dw(t),$$

where μ is the instantaneous drift per unit time, σ_A is the instantaneous volatility of $\mathcal{S}(t)$, and $w(t)$ is a standard Brownian motion (see Crack [2009] for introductory discussion of Brownian motions). Under our assumptions, and under the risk-neutral probability measure, the process is $d\mathcal{S}(t) = \sigma_A dw^*(t)$.

Assume a strike of $\mathcal{X}$. Note that under the risk-neutral probability measure with $r = 0$ the process $\mathcal{S}(t)$ is given by equation B.2:

$$\mathcal{S}(t) = \sigma_A w^*(t) \tag{B.2}$$

The call price is the discounted expected payoff under the risk-neutral probability measure, as follows:

$$\begin{aligned} c(t) &= e^{-r(T-t)}\, E^*[\max(\mathcal{S}(T) - \mathcal{X}, 0) \mid \mathcal{S}(t)] \\ &= E^*[\max(\mathcal{S}(T) - \mathcal{X}, 0) \mid \mathcal{S}(t)] \end{aligned}$$

From equation B.2, it follows that

$$\begin{aligned} \mathcal{S}(T) &= \mathcal{S}(t) + \sigma_A(w^*(T) - w^*(t)) \\ &= \mathcal{S}(t) + \sigma_A \mathcal{W}^*, \end{aligned}$$

where $\mathcal{W}^* \equiv w^*(T) - w^*(t)$ is Normal $\mathcal{N}(0, T-t)$ under the risk-neutral probability measure. Now let "v" play the part of $\mathcal{W}^*$ distributed as $\mathcal{N}(0, T-t)$. Then the call price is given by the following integration over the Normal density:

$$c(t) = \int_{v_0}^{+\infty} (\mathcal{S}(t) + \sigma_A v - \mathcal{X})\; f_V(v)\; dv,$$

where

$$f_V(v) = \frac{1}{\sqrt{2\pi}\,\sqrt{T-t}} e^{-\frac{1}{2}\left(\frac{v}{\sqrt{T-t}}\right)^2}$$

is the pdf of $v \sim \mathcal{N}(0, T-t)$, and

$$v_0 \equiv \frac{\mathcal{X} - \mathcal{S}(t)}{\sigma_A}$$

is the boundary value of v at which $(\mathcal{S}(t) + \sigma_A v - \mathcal{X})$ changes sign.

The remainder of the proof is left to the reader. The final result is

$$c(t) = \sigma_A \sqrt{T-t} \left\{ \frac{e^{-\frac{1}{2}d^2}}{\sqrt{2\,\pi}} + N(d)\, d \right\}, \tag{B.3}$$

$$\text{where } d = \frac{\mathcal{S}(t) - \mathcal{X}}{\sigma_A \sqrt{T-t}}.$$

The arithmetic Brownian motion pricing formula (equation B.3, above) is not well known. This is because an arithmetic Brownian motion is not a reasonable assumption for a price process: arithmetic Brownian motions can assume negative values. However, the geometric Brownian motion assumed by Black and Scholes (1973) is always non-negative, just as a price process should be. The importance of pricing options on stock catapulted the Black-Scholes formula (and the geometric Brownian motion beneath it) to super-stardom, while the pricing formula for the arithmetic Brownian motion languishes in relative obscurity.

Let's have a little history. Louis Jean Baptiste Alphonse Bachelier finished his mathematics PhD thesis at the Sorbonne in Paris in January 1900.[22] The topic of his thesis was the pricing of options contracts traded on the Paris Bourse.[23] Bachelier (1900) assumes that stock prices are Normally distributed and follow an arithmetic Brownian motion. He also assumes that expected returns on stocks (and on investments in general) are zero. Bachelier was the first to publish payoff diagrams for a European call option. Bachelier was also the first mathematician to use the "reflection principle."[24] Bachelier's derivation of the mathematical properties of Brownian motion predates by five years Albert Einstein's 1905 work on Brownian motion (Einstein [1905]). Bachelier even tested the predictions of his model using actual option prices on the Paris Bourse and found them not too far wrong.[25]

Unfortunately, Bachelier's assumptions violate some basic economic principles. In particular, he violates limited liability, time preference, and risk aversion (see Samuelson [1965, p13] for discussion). However, the significant contributions of Bachelier's thesis mean that he is rightfully considered the "father of modern option pricing theory" (Sullivan and Weithers [1991]).

In the special case when $\mathcal{S}(t) = \mathcal{X}$ (the option is struck at-the-money), equation B.3 reduces to[26]

$$c_A(t) = \sigma_A \sqrt{\frac{T-t}{2\pi}}, \tag{B.4}$$

[22]The Sorbonne was the prestigious University of Paris founded by Robert de Sorbon in 1253. The Sorbonne was split into 13 units during the period 1968–1970. Nowadays, the name "Sorbonne" refers to the original university or to three of the 13 units that retain the title as part of their name. I had the pleasure of visiting the Sorbonne as a tourist in both 1998 and 1999. It is on the Left Bank, not far from Notre Dame.

[23]A very brief look at Bachelier's model is in Appendix A of Smith (1976); a full translation appears in Cootner (1964). Note that my option pricing formula, equation B.3, is mathematically equivalent to equation A.5 in Smith (1976). On an historical point of some coincidence, note that as I write this (for my first edition in 1995) it is Louis Bachelier's 125^{th} birthday. Bachelier was born in Le Havre, France, on March 11th, 1870. See also Zimmerman and Hafner (2007) for a discussion of the Bronzin option pricing work of 1908.

[24]It is also known as the method of "reflected images." If you do not yet know what the reflection principle is, you probably do not need to know. If you are curious, see Harrison (1985, p7) for details.

[25]Samuelson (1973, Footnote 2, p6) compares Bachelier (1900) and Einstein (1905). He declares Bachelier dominant "in every element of the vector." See Samuelson (1973) for further discussion (and criticism) of Bachelier and other topics in the mathematics of speculative prices.

[26]This is equation 4.7 in Samuelson (1973).

where my "A" indicates that the underlying process, $\mathcal{S}(t)$, is an *arithmetic* Brownian motion, and σ_A is the standard deviation of the level of $\mathcal{S}(t)$.

Equation B.4 was derived assuming $r = 0$, and $\mathcal{S}(t) = \mathcal{X}$, plus the assumption of arithmetic Brownian motion. You might reasonably ask how does equation B.4 compare to Black-Scholes for an at-the-money call option when $r = 0$?

The Black-Scholes formula for pricing a standard European call on a non-dividend-paying stock reduces to equation B.5 in the special case when $r = 0$ and $S(t) = X$ (i.e., the option is struck at-the-money):

$$c_{BS}(t) = S(t)\left[N\left(+\frac{\sigma}{2}\sqrt{T-t}\right) - N\left(-\frac{\sigma}{2}\sqrt{T-t}\right)\right], \tag{B.5}$$

where $S(t)$ follows a *geometric* Brownian Motion, and σ is the standard deviation of continuously compounded returns on the stock price, $S(t)$.

When σ is small, equation B.5 may be approximated as[27]

$$c_{BS}(t) \approx S(t)\sigma\sqrt{\frac{T-t}{2\pi}}. \tag{B.6}$$

Compare equation B.6 with equation B.4. In the arithmetic Brownian motion case, σ_A is the standard deviation of the level of the price process $\mathcal{S}(t)$; in the geometric Brownian motion case, σ is the standard deviation of continuously compounded returns. Standard deviation of price is, however, approximately equal to price times the standard deviation of continuously compounded returns. It follows that the pricing in equations B.4 and B.6 is consistent, even though the first uses an arithmetic Brownian motion (supposedly incorrect), and the second uses a geometric Brownian motion. Thus, the Black-Scholes formula reduces to the century-old Bachelier formula.

In my other book *Basic Black-Scholes*, (Crack [2009]) I demonstrate the general ABM case where we assume neither that the option is at the money, nor that $r = 0$.[28] The formula for the call option price in this case is given by equations B.7–B.9. See Crack (2009) for full details of the derivation.

$$\begin{aligned} c(t) &= e^{-r(T-t)}\sigma_A\sqrt{\frac{e^{2r(T-t)}-1}{2r}}\left[N'(d) + N(d)\cdot d\right] && \text{(B.7)}\\ &= e^{-r(T-t)}\sigma_A\sqrt{\frac{e^{2r(T-t)}-1}{2r}}\left[\frac{e^{-\frac{1}{2}d^2}}{\sqrt{2\,\pi}} + N(d)\cdot d\right] && \text{(B.8)}\\ \text{where } d &= \frac{\mathcal{S}(t)e^{r(T-t)} - \mathcal{X}}{\sigma_A\sqrt{\frac{e^{2r(T-t)}-1}{2r}}}. && \text{(B.9)} \end{aligned}$$

[27]This approximation appears in Brenner and Subrahmanyam (1988). They use a Taylor series derivation, but less formally it follows because $[N(z) - N(-z)]$ is just the area under the Standard Normal pdf from $-z$ to z. With σ small, you can approximate the area by length times height. The length is $\sigma\sqrt{T-t}$; for small σ, the height is close to the height of the Standard Normal pdf at its peak: $\frac{1}{\sqrt{2\pi}}$ (recall that $\frac{1}{\sqrt{2\pi}} \approx 0.4$).

[28]I thank Mikhail Voropaev for contributing this idea. Any errors are mine.

Answer 2.20: Black-Scholes in your head!? This technique is so well known that some interviewers just ask if you can do it, and if you say yes they move on. It's not worth the gamble if you don't know it.[29]

Traders use the arithmetic Brownian motion approximation (or Black-Scholes reduced formula) from Answer 2.19 as a rough but fundamental call pricing relationship:

$$c(t) \approx \sigma S\sqrt{\frac{T-t}{2\pi}}, \tag{B.10}$$

where σ is the standard deviation of returns or where σS is replaced by the standard deviation of prices. You should also note that this versatile little formula prices *both* puts and calls. Why is this? Well, if interest rates are low, and the option is struck at-the-money, then in the absence of dividends a call and put have the same value—just use put-call parity.

Many times interviewees are asked to price an option in their head where the interest rate is zero and the option is struck at-the-money. You should, therefore, know that the option pricing formulae of both Black-Scholes and Bachelier reduce to equation B.10 and that it works for both for puts and calls. I expect you to be able to evaluate equation B.10 in your head in less than 10 seconds if asked to in an interview. How can you do this so quickly? Well, $\frac{1}{\sqrt{2\pi}} \approx 0.4$, and for three months, six months, or one year to maturity, you have $\sqrt{0.25} = 0.50$, $\sqrt{0.50} \approx 0.70$, and $\sqrt{1} = 1$, respectively. Of course, it helps that they usually give you easy numbers. For example, if $S = \$100$, $\sigma = 0.40$, and $(T - t) = 0.25$, the formula gives \$8 $(0.4 \times 0.4 \times 100 \times 0.5)$ whereas Black-Scholes proper gives \$7.97—not bad at all! The approximation is usually accurate to within a couple of percentage points.

Answer 2.21: This is a common type of question requiring fundamental knowledge. The only thing that changes between the two options is the time until expiration. The important knowledge here is how the value of a call changes with time to expiration.

You should remember that in the special case where $r = 0$, and the option is struck at-the-money [so that $S(t) = X$], the Black-Scholes European call option pricing formula may be approximated by the following (see discussion in Answer 2.20):

$$c(t) = \sigma_A\sqrt{\frac{T-t}{2\pi}},$$

where σ_A is not standard deviation of continuously compounded returns, but standard deviation of price. From this approximation, you can see that if the call is at-the-money, the call value increases at something like the square root of the term to maturity (if you double term to maturity, value increases by 40% to 50%). You must be very comfortable with this approximation.

[29] See Haug (2001) for related material.

The above approximation is a good place to start. However, a full answer recognizes that the response of call value to term to maturity depends heavily upon whether the call is in-the-money, at-the-money, or out-of-the money. Sensitivity to term to maturity decreases as you move into-the-money, down to zero in the limit if you are very deep in-the-money. Sensitivity to term to maturity increases as you move out-of-the-money. Doubling the term to maturity can easily double, triple, or quadruple the value of the call if it is well out-of-the-money. The effect is greater the further out-of-the-money the call is—this is why deep out-of-the-money options are sometimes called "lottery tickets."

You can see this effect clearly if you compare the prices of actively traded equity options and LEAPS.[30] Go to the listed options quotations in the third section of the *Wall Street Journal*. Choose a stock for which both equity options and LEAPS are traded. Compare the prices of call options on your chosen stock that have the same strike, but different terms to maturity (e.g., three months, six months, one year, and two years). If you do this comparison for different strike prices, you should see that an extension in term to maturity has the most impact on call option prices when the options are out-of-the-money. You should see that for call options that are deep out-of-the-money, doubling the term to maturity can easily quadruple the value of the call option.

In simple terms, if you extend the term to maturity, the option has more opportunities to finish in-the-money, the present value of the cost of exercising decreases, and the call value increases. The increase in the call value depends upon the initial likelihood that the call will finish in-the-money. This likelihood is small if the option is well out-of-the-money. Thus an increase in term to maturity produces a proportionately greater increase in the value of an option that is out-of-the-money.[31]

A caveat. The approximation formula $c(t) = \sigma\sqrt{\frac{T-t}{2\pi}}$ prices at-the-money European-style puts and calls when $r = 0$. However, it has its limitations. For example, if $r \neq 0$, then the value of a deep in-the-money European put *decreases* as time to maturity extends. If the put is deep in-the-money, then life is already as good as it gets (the put has limited upside). You want to exercise now and take the money. Extending the life of the option pushes the expected benefit further away and decreases the put's value.

Answer 2.22: I give two different methods for answering this question. If the standard deviation is \$20, not \$10, then double the answers given.

[30] LEAPS are "Long-term Equity AnticiPation Securities." That is, LEAPS are long-term options. LEAPS have terms to maturity of up to three years. The term to maturity of standard exchange-traded equity options does not exceed eight months. LEAPS are not exotic options, but exchange-traded standardized options contracts. Like standard equity options, all LEAPS are American-style options. Unlike standard equity options, equity LEAPS all expire in January; index LEAPS all expire in December (Options Clearing Corporation [1993]).

[31] For a very helpful practitioner view on the interpretation of partial derivatives of call price with respect to each option pricing parameter, see Chance (1994).

FIRST SOLUTION
As a loose rule of thumb, the standard deviation of price per period ($10 here) is a rough measure of the average possible upside move or downside move in stock price over the next period. You have approximately half-a-chance of finishing in-the-money, and half-a-chance of finishing out-of-the-money (or "under water," as it is sometimes called). The expected payoff is, therefore, roughly $\left(\frac{1}{2} \times \$0\right) + \left(\frac{1}{2} \times \$10\right) = \$5$. In fact, the shape of the Lognormal distribution of final stock price means that the expected payoff is slightly less than $5 (it is around $4).

SECOND SOLUTION
In the case where $r = 0$, and the option is struck at-the-money [so that $S(t) = X$], the Black-Scholes option pricing formula may be approximated by the following (see discussion in Answer 2.20):

$$c(t) = \sigma_A \sqrt{\frac{T-t}{2\pi}},$$

where σ_A is not standard deviation of continuously compounded returns, but standard deviation of price. With $T - t = 1$, and $\frac{1}{\sqrt{2\pi}} \approx 0.40$ (memorize that one), the standard deviation of price of $10 implies a call price of around $4. Note that this technique is more accurate than the first, giving $4 instead of $5.

Answer 2.23: The answer cannot be found exactly in the Black-Scholes framework, but you can get a good estimate.[32] Increasing the implied volatility σ by 25% (from 0.20 to 0.25) on one day out of 100 in the option's life is the same (to a first-order approximation) as increasing σ^2 by 50% on one day out of 100 in the option's life.[33] This averages out to something like increasing σ^2 by 0.5% for every day remaining in the option's life (i.e., multiplying the average σ^2 by a factor 1.005).[34] Using the approximation (see page 128) $c(t) \approx \sigma S(t)\sqrt{\frac{T-t}{2\pi}}$, we see that multiplying σ^2 by M has the same effect on $c(t)$ as multiplying $T-t$ by M. This is because each of σ^2 and $(T-t)$ appear in the option formula under a square root sign—either implicitly or explicitly. It follows that multiplying σ^2 by a factor 1.005 is equivalent to increasing the term to maturity by something like 0.5% (i.e., one-half of a day for a 100-day option). That is, increasing σ^2 by 50% on one day is equivalent to increasing the length of one day by 50%.

Either of the adjustments mentioned increases the value of an at-the-money option by a factor of about $\sqrt{M}$—a quarter of a percent here. Note that the equivalence of the 50% increase in σ^2 on one day and the extension of option life by half a day is a general result—because variance is linear in time. However, the conclusion that either of these adjustments increases option value by about a quarter of a percent applies only to at-the-money options. If an option is deep in-the-money, the adjustments mentioned may have little or no effect on option value; if an option is deep out-of-the-money, the adjustments mentioned may increase the option value by substantially more than a quarter of a percent.

[32] Francis Longstaff suggested to me that an important option pricing problem is the handling of "event risk" (personal communication [September 25, 1998]). For example, how do you price a 14-day option on a stock whose CEO is scheduled to make an important announcement in seven days. I think this question is a loose attempt at this issue.

[33] The "implied volatility" is the volatility figure implicit within an option price, assuming that market participants value options using the Black-Scholes formula. The "implied vol" appears first in the literature in Latané and Rendleman (1976).

[34] As an aside, note this for the Black-Scholes formula: If we increase the calendar term to maturity, but still call it "one period," then we need to increase σ. However, if we increase $(T-t)$ ("term to maturity" or "the number of periods") without changing the length of one period in the model, we do not need to change σ.

Answer 2.24: A give-away question! A long straddle is a long call plus a long put with the same strike. If you hold the straddle until maturity, then you need a price change of more than $5 either way in the underlying to profit. A smaller price change, however, can lead to profits if it happens before maturity. For example, using Black-Scholes (ignoring that CBOE equity options are American-style), if $\sigma = 0.357$, $T - t = 0.5$, $S = \$25$, and $r = 0.02$, then a straddle struck at $25 costs $5. If the price of the underlying suddenly jumps to $27, then the straddle is suddenly worth $5.50 and you have an immediate 10% gain. See table B.1 for details.

Table B.1: Straddle Prices when the Stock Price Jumps

	Stock Price = $25.00	Stock Price = $27.00
Price of the Call (X = $25)	$2.625	$3.875
Price of the Put (X = $25)	$2.376	$1.626
Price of the Straddle (sum)	$5.001	$5.502

Note: The option prices in the table are calculated using volatility of $\sigma = 0.357$ per annum, time to maturity of $T - t = 0.5$ years, a riskless rate of $r = 0.02$ per annum, and the Black-Scholes formula. A long straddle is a long call plus a long put with the same strike. A straddle struck at $25 costs $5 when stock price is $S = \$25$, but if the stock price jumps immediately to $27, the straddle is worth $5.50, giving an immediate 10% gain, ignoring transactions costs.

Answer 2.25: The Eurodollar futures contract is the most popular short-term interest rate futures contract. The contract value used for marking-to-market at the end of the day is $\$10,000 \times [100 - \frac{90}{360}\delta]$, where δ is the settlement discount rate. Between settlements, the market participants determine, through supply and demand, what is considered a fair discount. At maturity, the discount δ must converge to the three-month LIBOR US dollar rate. Note that if the discount is 5%, then $\delta = 5.0$ in the above calculation, not 0.05.

The three-month LIBOR rate is typically about 40 to 50 bps (i.e., 0.40 to 0.50 percentage points) higher than the yield on three-month treasury bills (this compensates for default risk of London banks).[35] The discount rate δ is thus highly correlated with US interest rates. The contract value is highly negatively correlated with δ, and thus highly negatively correlated with US interest rates.

Suppose that you are long a Eurodollar future. If US interest rates rise, the contract value declines, and you finance your loss at a relatively high rate. If US interest rates fall, then the contract value rises, but you invest the marked-to-market gains at relatively low rates. If you hold the forward, rather than the future, you do not have day-to-day gains and losses, so you are not hurt in the same way by these opportunity costs. Other things being equal, you would rather have the Eurodollar forward contract than the Eurodollar futures contract. If the discounts are the same (as stated), then there is a mis-pricing.[36] With the current mis-pricing, I would choose to go long the Eurodollar forward and short the Eurodollar future.

Answer 2.26: It makes much more sense to simulate the underlying and find the payoffs to the call, than it does to simulate the process for the call itself. It is difficult to model the call, because the instantaneous volatility of the call changes whenever the leverage of the call changes (assuming the underlying is of constant volatility). The leverage of the call changes whenever the stock price moves (and it even changes if the stock price does not move—simply because of time decay).

Answer 2.27: I think a quick review of "mortgage-backs" is in order before addressing the question. Mortgage-backed securities are shares in portfolios of mortgages. The value of all US agency mortgage-backed securities outstanding was around \$8.9 trillion as of late-2008—that is, \$8,900,000,000,000 (according to SIFMA's web site).

Owners of mortgage-backs are exposed to "prepayment risk," and "extension risk." Prepayment risk is the risk that interest rates will fall, and borrowers will exercise their right to refinance at lower rates (they exercise their call option on the mortgage). The problem is that the holders of the mortgage-backs, therefore, get repaid when interest rates are low—the worst possible time to receive the money. Conversely, extension risk is the risk that interest

[35]This is an estimate only. There is tremendous variation in the spread. The "Ted Spread" is the Eurodollar futures less T-bill futures index point spread (with same delivery month). The Ted Spread was a typical 45.9 bps on June 15, 2009. During the Global Credit Crisis, however, it spiked up to over ten times this level—reaching 465 bps on October 10, 2008—as credit risk fears pushed three-month USD LIBOR up to 4.82% and demand for safety pushed three-month T-bill yields down to 24 bps.

[36]If the underlying were strongly *positively* correlated with US interest rates, then the futures contract would be more attractive than the forward. This is because daily gains can be invested at relatively high rates, while daily losses are financed at relatively low rates (see Hull [1997, pp55–56] for more details).

rates will rise, and borrowers will slow down their rate of repayment—meaning that holders of mortgage-backs get fewer dollars to invest at precisely the best time for them to be investing. Mortgage-back investors thrive when interest rate volatility is low.

The simplest mortgage-back is a "pass-through"—each share in the mortgage pool provides a prorata share in the cash flows to the pool, and thus each share has identical risk and return characteristics.

Collateralized mortgage obligations (CMO's) are a type of mortgage-back that splits the mortgage pool up into "tranches" (the French word for "slice"). Unlike a pass-through, which gives equal shares to all holders, the tranches are unequal shares. Take a simple example with only four tranches: "A," "B," "C," and "Z." The A, B, and C shares all receive regular coupons. The A shares are retired (i.e., the principal is repaid) ahead of the other tranches by using the earliest prepayments by borrowers. The B shares are retired, through prepayments, only after the A shares are gone. The C shares are retired, through further prepayments by borrowers, only after the A and B shares are gone. The Z shares receive no payouts whatsoever until all of the A, B, and C shares are gone. You may think of the Z shares as being like zero-coupon bonds with a life equal to the life of the longest-lived mortgages in the pool. CMO tranches thus provide *different* risk-return profiles—in contrast to pass-throughs.

With borrowers long a call on the mortgage (i.e., the right to buy back the mortgage by prepayment), holders of mortgage-backs are short a share of each of these calls from the mortgage pool. You will recall, of course, that long calls have positive convexity and that short calls have negative convexity.

In the absence of the call feature of a mortgage-back (the fact that borrowers have the right to prepay early), the mortgage back has positive convexity as a function of interest rates—just like an ordinary non-callable bond (Sundaresan [1997, p393]). However, when interest rates are low, the call feature becomes important to borrowers. If interest rates fall, all borrowers will refinance by the time rates have fallen to some critical value. At this stage, the mortgage-back is worth par. When interest rates are low, the importance of the call feature (a short call position to the mortgage-back holder) means that the mortgage-back can acquire negative convexity.[37] Negative convexity is also called "compression to par" because of the convergence of the security's value to par as interest rates fall (Sundaresan [1997, p394]).

Note that although the mortgage-back value may have negative convexity, it is still downward sloping as a function of interest rates. However, if interest rates are low and close to the coupon rate of the mortgage back, then an increase in volatility of interest rates can decrease the value of the security (Sundaresan [1997, p394]). This result follows because the holder of the mortgage back is

[37]In fact, this is a feature of any callable bond—if interest rates fall far enough, the call feature kicks in and imparts negative convexity to the security.

short the calls that the borrowers are long—and calls increase in value with volatility.

Now to the interview question. If you are long mortgage-backs, and you expect a bond market rally, then you expect bond yields to fall and bond prices to rise. Thus, your position will gain in value. The question is, which sign on convexity would maximize the gain (+ or −)? Positive convexity provides a steeper downward sloping plot of security value as a function of interest rates, and this in turn implies a larger gain if rates fall—thus, you prefer positive convexity.[38]

A full answer notes that we have assumed a *parallel* shift in the yield curve. If the yield curve steepens or flattens, the answer could change. Whether additional convexity helps or hurts you depends upon the type of yield-curve shift and the particular bonds under consideration. It needs to be evaluated on a case-by-case basis. See the related discussion beginning on page 179.

Answer 2.28: The hedging strategy is naive. This is called a "stop-loss strategy" (Hull [1997, p310]). At first glance, it replicates the payoff to the call. However, purchases and sales cannot be made at the strike price. When the stock is near the strike, you cannot know whether it will cross over the strike price or not. You have to wait until the stock price crosses the strike price. This means you end up making purchases at a price slightly higher than the strike and sales at a price slightly lower than the strike. The closer to the strike you try to time your trades, the more frequently you can expect to have to trade. You can get eaten alive by transactions costs (see Hull [1997, p310]).

A second criticism is that the timing of the cash flows to the option and the hedge are different—it is not a hedge (see Hull [1997, p310]).

Answer 2.29: This question and the next are the most popular stochastic calculus interview questions. Although this is ostensibly a stochastic calculus question, the answer relies only upon Riemann calculus. If you were stuck and looking for a hint, then maybe this is enough to get you going.[39]

Let $I_T(\omega)$ denote the integral $\int_0^T w(t,\omega)dt$. In this integral, t measures time along sample paths, and ω is an element of the sample space Ω (i.e., ω corresponds to a particular possible sample path). Since $w(t)$ has continuous paths with probability one (i.e., for almost every $\omega \in \Omega$ the path is continuous), this integral is a Riemann integral evaluated pathwise for any fixed $\omega \in \Omega$. The

[38] Convexity is not such an issue if you expect a bond market rout. When prices fall and rates rise, prepayment becomes less attractive, and the call option in the hands of the borrowers assumes less importance—and so does the negative convexity the call is able to impart to your mortgage-back security value.

[39] I thank Taras Klymchuk and Ian Short for suggesting related solution techniques. I am responsible for any errors.

Riemann integral is just (in its simplest form)

$$I_T = \lim_{n\to\infty} S_n, \text{ where } S_n \equiv \sum_{i=1}^{n} (t_i - t_{i-1})\, w(t_{i-1}), \text{ and}$$

$$t_i \equiv \frac{Ti}{n}, \quad 0 = t_0 < t_1 < \ldots < t_n = T.$$

We may rearrange terms as follows:

$$\begin{aligned}
S_n &= \sum_{i=1}^{n} (t_i - t_{i-1})\, w(t_{i-1}) \\
&= (t_1 - t_0)w(t_0) + (t_2 - t_1)w(t_1) + \ldots + (t_n - t_{n-1})w(t_{n-1}) \\
&= -t_0 w(t_0) + t_1\left[w(t_0) - w(t_1)\right] + t_2\left[w(t_1) - w(t_2)\right] \\
&\qquad + \ldots + t_{n-1}\left[w(t_{n-2}) - w(t_{n-1})\right] + t_n w(t_{n-1}) \\
&= -t_0 w(t_0) + \sum_{i=1}^{n-1} t_i\left[w(t_{i-1}) - w(t_i)\right] \\
&\qquad + t_n \sum_{i=1}^{n-1}\left[w(t_i) - w(t_{i-1})\right] + t_n w(t_0) \quad \text{(a telescoping series)} \\
&= \sum_{i=1}^{n-1} (t_n - t_i)\left[w(t_i) - w(t_{i-1})\right], \quad \text{a.e.}
\end{aligned}$$

The last line follows because $w(t_0) \equiv w(0) = 0$ a.e. (i.e. almost everywhere) by definition. So, S_n is just a weighted sum of increments of a standard Brownian motion. It is well known that such increments are independently Normally distributed and that a finite sum of constant-weighted independent Normals is also Normal. Thus, S_n is Normal for each n. Having established Normality, we can deduce the mean and variance using algebra or calculus; I will do both.

> **Advice:** *Keep your hands off your face during an interview! About one in three people I interview stick their fingers up their nose while I am talking to them. At the end of the interview I do not want to shake your hand and I do not want to hire you.*

ALGEBRA DERIVATION OF MEAN AND VARIANCE

Now, notice that $(t_n - t_i) = Tn/n - Ti/n = T(n-i)/n$, and $(T(n-i)/n)[w(t_i) - w(t_{i-1})] \sim \mathcal{N}(0, (n-i)^2T^3/n^3)$ (because $w(t_i) - w(t_{i-1}) \sim \mathcal{N}(0, T/n)$). A little bit of algebra gives us that $\sum_{i=1}^{n-1}(n-i)^2 = \sum_{i=1}^{n-1} i^2 = n(n-1)(2n-1)/6$ (using Answer 1.41), so that $\sum(n-i)^2T^3/n^3 = n(n-1)(2n-1)T^3/(6n^3) \to T^3/3$ as $n \to \infty$. That is, conditional on time 0 information, $I_T(\omega)$ is distributed as $\mathcal{N}(0, T^3/3)$.

CALCULUS DERIVATION OF MEAN AND VARIANCE
The mean is just $E(I_T) = \int_0^T E[w(t)]dt = 0$. With a mean of zero, the variance is just the second non-central moment

$$\begin{aligned} V(I_T) &= E(I_T^2) \\ &= E\left\{\left(\int_0^T w(t)dt\right)\left(\int_0^T w(s)ds\right)\right\} \\ &= \int_0^T\int_0^T E\left[w(t)w(s)\right]dtds. \end{aligned}$$

Now, you will recall that $w(t)$ is a process with independent increments. Let us assume, for the moment, without loss of generality, that $s < t$. Then, $w(t) = w(s) + (w(t) - w(s))$, and $w(t)w(s) = w^2(s) + (w(t) - w(s))w(s)$. It follows that

$$E\left[w(t)w(s)\right] = E\left[w^2(s)\right] = s,$$

using independent increments and the fact that $w(s)$ has a variance of s. More generally, $E\left[w(t)w(s)\right] = \min(t,s)$. Thus, we may write the variance of the integral as follows:

$$\begin{aligned} V(I_T) &= \int_0^T\int_0^T E\left[w(t)w(s)\right]dtds \\ &= \int_0^T\int_0^T \min(t,s)dtds \\ &= \int_0^T\left(\int_0^s tdt + \int_s^T sdt\right)ds \\ &= \int_0^T\left(\frac{s^2}{2} + s(T-s)\right)ds \\ &= \int_0^T\left(sT - \frac{s^2}{2}\right)ds \;=\; \left(\frac{T^3}{2} - \frac{T^3}{6}\right) = \frac{T^3}{3}, \end{aligned}$$

demonstrating again that $I_T(\omega)$ is distributed as $\mathcal{N}(0, T^3/3)$.

Answer 2.30: Do you need a hint? This problem requires Itô's Lemma and not much else. Now go back to the problem and stop peeking at the solutions.[40]

[40] I had the pleasure of attending a symposium in honor of Norbert Wiener at MIT in October 1994 ("The Legacy of Norbert Wiener: A Centennial Symposium"). Two seats to my left sat Kiyoshi Itô—of Itô's Lemma fame. Although 79 at the time, Professor Itô did not appear old. He was of small build and very distinguished looking. He spoke clearly in somewhat halting English, and his good-natured humor was infectious. Previous and future Nobel Prize winners, respectively, Paul Samuelson and Robert Merton also spoke, and it seems that Itô's Lemma was in fact a footnote in a paper of Itô's. They joked that it should be called "Itô's Footnote" instead—but that does not have the same ring to it. In 2006, at age 91, Itô was awarded the Gauss Prize in Mathematics (Protter [2007]).

If we apply Itô's Lemma to $F(t,w) \equiv \frac{w^2(t)}{2}$, we find[41]

$$dF = F_t dt + F_w dw + \frac{1}{2}F_{ww}(dw)^2 = w(t)dw(t) + \frac{1}{2}dt.$$

This notation means precisely

$$F(T) - F(0) = \int_0^T w(t)dw(t) + \frac{1}{2}\int_0^T dt = \int_0^T w(t)dw(t) + \frac{T}{2}.$$

Given the definition of $F(t)$, it follows immediately that

$$\int_0^T w(t)dw(t) = \frac{w^2(T) - T}{2} \quad \text{a.e.}$$

It should be noted that the expected value of the right-hand side of the equality is zero. This is consistent with the expected value of the left-hand side of the equality being zero also.

Answer 2.31: Most people incorrectly deduce that the derivative security is worth \$1. If you got this answer, go back right now and think some more. I present three solution techniques: the first uses standard no-arbitrage arguments; the second uses partial differential equations (PDE's); the third uses stopping times.[42]

FIRST SOLUTION
You are an investment banker. Assume there exists a derivative security that promises one dollar when IBM hits \$100 for the first time. If this security is marketable at *more* than \$0.75, then you should issue 100 of them and use \$75 of the proceeds to buy one share of IBM. If IBM ever hits \$100, sell the stock and pay \$1 to each security holder as contracted. You sell the securities, perfectly hedge them, and still have money in your pocket. By no-arbitrage, the security cannot sell for more than \$0.75.

The converse is that if this security costs *less* than \$0.75, you should buy 100 of these securities financed by a short position in one IBM share. For this to establish \$0.75 as a lower bound on the security price (and, therefore, to pinpoint the price at \$0.75—the solution given to the interviewee by the Wall Street firm), you need to assume that you can roll over a short position *indefinitely*. This assumption seems reasonable for moderate amounts of capital. However, it is not clear to me that this is a reasonable interpretation of "ignore any short sale restrictions" when larger quantities of capital are involved. As one colleague said to me: "If it were possible to short forever, I'd short stocks with face value of a billion dollars, consume the billion, and roll over my short position forever." This seems to be an arbitrage opportunity.

[41] I thank Taras Klymchuk for suggesting this solution technique. I am responsible for any errors.
[42] If you want a good introductory book on PDE's, I recommend Farlow (1993). I loved this book when I was a student. I still find it a breath of fresh air compared to my other math books.

We conclude that \$0.75 is a clear upper bound by no-arbitrage, and thus \$1 cannot be the correct answer. Whether or not \$0.75 is also a lower bound is arguable (but it seems to make sense for moderate amounts of capital). The second solution technique also establishes \$0.75 as the value of the security.

SECOND SOLUTION

This technique is more advanced and may be beyond the average candidate.[43] The derivative value V must satisfy the Black-Scholes PDE (Wilmott et al. [1993]):

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

The boundary conditions that make sense for $V(S,t)$ are

$$\begin{aligned} V(S=100, \text{ any } s>t) &= \$1, \text{ and} \\ V(S=0, \text{ any } s>t) &= \$0. \end{aligned}$$

Let us simplify our lives by searching initially for a solution that is affine in S: $V(S,t) = kS(t) + l$, for some constants k and l.[44]

The two boundary conditions imply that

$$\begin{aligned} k \times \$100 + l &= \$1, \text{ and} \\ k \times \$0 + l &= \$0. \end{aligned}$$

From these we deduce that $k = \frac{1}{100}$, and $l = 0$. The functional form $V(S,t) = \frac{1}{100}S(t)$ satisfies the Black-Scholes PDE and the two boundary conditions and is thus the derivative value. In the special case where $S(t) = \$75$, we get $V = \$0.75$, as for the first technique.

THIRD SOLUTION

Following Shreve (2004, p297), we will take a "first passage" or "stopping time" approach. This technique is even more advanced than the previous one.[45]

Let now be time $t = 0$ and consider a derivative with lifespan T that pays \$1 if stock price $S(t)$ hits a barrier at level $B > S(0)$ before time T. In our case $B = \$100$ and $S(0) = \$75$. In the Black-Scholes world we have risk-neutral stochastic differential equation $dS(t) = rSdt + \sigma S(t)d\widetilde{W}(t)$ with solution

$$S(t) = S(0)e^{(r-\frac{1}{2}\sigma^2)t+\sigma\widetilde{W}} = S(0)e^{\sigma\widehat{W}},$$

[43] I thank Alan J. Marcus for suggesting this type of solution technique. I am responsible for any errors.

[44] An affine function involves both a linear portion, kS, and a constant, l. On a two-dimensional plot, a linear function goes through the origin; whereas an affine function may have a non-zero intercept.

[45] I thank Olaf Torne for suggesting this type of solution technique. I am responsible for any errors.

where $\widetilde{W}(t)$ is a standard Brownian motion, $\widehat{W}(t) \equiv \alpha t + \widetilde{W}(t)$, and $\alpha \equiv \frac{1}{\sigma}\left(r - \frac{1}{2}\sigma^2\right)$.

Let $\widehat{M}(T) \equiv \max_{0\le t\le T}\widehat{W}(t)$, then $\max_{0\le t\le T} S(t) = S(0)e^{\sigma\widehat{M}(t)}$. The stock price hits the barrier B before time T if and only if this maximum stock price is larger than B. That is, $S(0)e^{\sigma\widehat{M}(t)} > B$. Simple algebra shows us that

$$\begin{aligned} \widetilde{P}\left(\max_{0\le t\le T} S(t) > B\right) &= \widetilde{P}(S(0)\,e^{\sigma\widehat{M}(t)} > B) \\ &= 1 - \widetilde{P}(S(0)\,e^{\sigma\widehat{M}(t)} \le B) \\ &= 1 - \widetilde{P}\left(\widehat{M}(t) \le \frac{1}{\sigma}\ln\left(\frac{B}{S(0)}\right)\right) \\ &= 1 - \widetilde{P}(\widehat{M}(t) \le b), \end{aligned}$$

where $b \equiv \frac{1}{\sigma}\ln\left(\frac{B}{S(0)}\right)$. We can use Shreve (2004, Corollary 7.2.2, p297) to deduce immediately that

$$\widetilde{P}(\widehat{M}(t) \le b) = N\left(\frac{b-\alpha T}{\sqrt{T}}\right) - e^{2\alpha b}N\left(\frac{-b-\alpha T}{\sqrt{T}}\right), \ b \ge 0.$$

It follows that the risk-neutral probability that we get our \$1 is[46]

$$\begin{aligned} &\widetilde{P}\left(\max_{0\le t\le T} S(t) > B\right) = \\ &= 1 - \left[N\left(\frac{b-\alpha T}{\sqrt{T}}\right) - e^{2\alpha b}N\left(\frac{-b-\alpha T}{\sqrt{T}}\right)\right] \\ &= N\left(\frac{-b+\alpha T}{\sqrt{T}}\right) + \left(\frac{B}{S(0)}\right)^{\left(\frac{2r}{\sigma^2}-1\right)} N\left(\frac{-b-\alpha T}{\sqrt{T}}\right) \\ &= \left[N\left(\frac{\ln\left(\frac{S(0)}{B}\right) + (r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)\right. \\ &\qquad \left. + \left(\frac{S(0)}{B}\right)^{\left(1-\frac{2r}{\sigma^2}\right)} N\left(\frac{\ln\left(\frac{S(0)}{B}\right) - (r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)\right] \quad (*) \end{aligned}$$

Equation $(*)$ gives the risk-neutral probability of the stock price touching the barrier and generating the payoff. If we multiply this probability by the \$1 payoff, we get the risk-neutral expected *future* payoff. If the interest rate is zero, then we do not need to discount, and plugging $r = 0$ into equation $(*)$ gives the value of the finitely-lived instrument that pays \$1 when the stock price hits B. If we take the limit as $T \to \infty$, the first $N(\cdot)$ term $\to 0$ (because its argument goes to $-\infty$), and the second $N(\cdot)$ term $\to 1$ (because

[46] I used $\ln(A) = -\ln\left(\frac{1}{A}\right)$ for $A > 0$, $1 - N(x) = N(-x)$, $A^x = \left(\frac{1}{A}\right)^{-x}$, and $e^{[g\ln(h)]} = h^g$.

its argument goes to $+\infty$). We are left with $\$1 \cdot \left(\frac{S(0)}{B}\right)$, which is just \$0.75 in our case.

From equation $(*)$, you should be able to deduce the risk-neutral probability that the stock price eventually hits the barrier as

$$\lim_{T\to\infty} \widetilde{P}(\text{hit B}) = \begin{cases} 1 & ; \text{ if } r \geq \frac{1}{2}\sigma^2, \\ \left(\frac{S(0)}{B}\right)^{\left(1-\frac{2r}{\sigma^2}\right)} & ; \text{ if } 0 \leq r < \frac{1}{2}\sigma^2. \end{cases}$$

Extension: Some candidates have very recently been asked this question in the case $r > 0$. My first two solutions do not rely upon $r = 0$, so the answer must be the same. Can we demonstrate this using this third solution technique? Well, to discount the expected future value at rate r we need to know how long to discount for. Let τ be the time at which $S(t)$ first hits B. This "first passage time" is distributed Inverse Gaussian with the following pdf:[47]

$$f(\tau) = \frac{\ln\left(\frac{B}{S(0)}\right)}{\sigma\sqrt{2\pi\tau^3}} e^{-\frac{1}{2}\left(\frac{\ln\left(\frac{B}{S(0)}\right)-(r-\frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}\right)^2} \quad \text{where} \quad \tau \geq 0$$

The general functional form of the Inverse Gaussian is

$$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}}\, e^{-\frac{\lambda}{2}\left(\frac{x-\mu}{\mu\sqrt{x}}\right)^2} \quad \text{where} \quad x \geq 0,\ \lambda > 0,\ \mu > 0.$$

In our case $x = \tau$, $\lambda = \frac{\left[\ln\left(\frac{B}{S(0)}\right)\right]^2}{\sigma^2}$ and $\mu = \frac{\ln\left(\frac{B}{S(0)}\right)}{\left(r-\frac{1}{2}\sigma^2\right)}$. With $B > S(0)$, the condition $\mu > 0$ implies that $r > \frac{1}{2}\sigma^2$, else we do not have a valid pdf for $f(\tau)$. If we proceed assuming $r > \frac{1}{2}\sigma^2$, our discount factor is the expected value of $e^{-r(\tau)}$ with respect to $f(\tau)$ such that $\tau \leq T$:

$$E[e^{-r(\tau)} I_{\{\tau \leq T\}}(\tau)],$$

where

$$I_X(x) \equiv \begin{cases} 1; & \text{if } x \in X, \\ 0; & \text{otherwise.} \end{cases}$$

So, we need to find the following discounted expected payoff:

$$\lim_{T\to\infty} \left\{ E[e^{-r(\tau)} I_{\{\tau\leq T\}}(\tau)] \cdot \$1 \cdot \left[N\left(\frac{\ln\left(\frac{S(0)}{B}\right) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) + \left(\frac{S(0)}{B}\right)^{\left(1-\frac{2r}{\sigma^2}\right)} N\left(\frac{\ln\left(\frac{S(0)}{B}\right) - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)\right]\right\}$$

[47] Given time taken, the distance traveled by a Wiener process is Gaussian, but given distance traveled, the time taken is Inverse Gaussian. So, the distribution of the first passage time for $\ln S(t)$ to hit $\ln B$ is also Inverse Gaussian (Crack [1998]).

For $r > \frac{1}{2}\sigma^2$, everything is surprisingly well behaved as $T \to \infty$. The discounting term $E[e^{-r(\tau)} I_{\{\tau\leq T\}}(\tau)] \to E[e^{-r(\tau)}]$, and it takes a page of tedious yet simple algebra to show that $E[e^{-r(\tau)}] = \frac{S(0)}{B}$ (hint: complete the square in the exponent). As we take the limit $T \to \infty$, the first $N(\cdot)$ term $\to 1$ (because its argument goes to $+\infty$), and the second $N(\cdot)$ term $\to 0$ (because its argument goes to $-\infty$). We are left again with $\$1 \cdot \left(\frac{S(0)}{B}\right)$, which is just \$0.75 as before. I find it interesting that in the two cases $r = 0$ and $r > \frac{1}{2}\sigma^2$ the limiting behaviour of the two cumulative Normal terms is opposite, but with the same end result. In the case $0 < r \leq \frac{1}{2}\sigma^2$, the instrument's value is the same, but I do not have a formal proof using Shreve's notation.

Finally, note that regardless of the value of r, the value of the finitely-lived instrument depends upon volatility, but the value of the perpetual instrument is independent of volatility.

Answer 2.32: There are two ways to proceed: the first way is to work out the pricing formula from first principles; the second way is to use Black-Scholes option pricing as it stands and make some ad hoc adjustments to it to account for the power payoff.

FIRST SOLUTION

I was unable to find a published pricing formula for the power call (with payoff $\max[S^\alpha - X, 0]$) or for the power put (with payoff $\max[X - S^\alpha, 0]$), so I followed a straight discounted expected payoff approach under risk-neutral probabilities. It is relatively straightforward to show that the value at time t of European power call and put options maturing at time T is

$$\begin{aligned} c(t) &= S^\alpha(t)e^{m(T-t)}N(d_1') - e^{-r(T-t)}XN(d_2'), \quad \text{and} \\ p(t) &= e^{-r(T-t)}XN(-d_2') - S^\alpha(t)e^{m(T-t)}N(-d_1'), \quad \text{where} \\ d_1' &= \frac{\ln\left(\frac{S(t)}{K}\right) + (r + (\alpha - \frac{1}{2})\sigma^2)(T-t)}{\sigma\sqrt{T-t}}, \\ d_2' &= \frac{\ln\left(\frac{S(t)}{K}\right) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} = d_1' - \alpha\sigma\sqrt{T-t}, \\ K &\equiv X^{\frac{1}{\alpha}}, \quad \text{and} \quad m \equiv \left(r + \frac{\alpha}{2}\sigma^2\right)(\alpha - 1). \end{aligned}$$

In the case $\alpha = 1$, the power option pricing formulae reduce to the standard Black-Scholes call and put pricing formulae.

The "delta" of the power call can be found by differentiating the power call pricing formula with respect to $S(t)$. The delta for the power call is given by

$$\begin{aligned}\Delta_{power\ call} &\equiv \frac{\partial c(t)}{\partial S(t)}\\ &= \alpha S^{(\alpha-1)}e^{m(T-t)}N(d_1')\\ &+\frac{X^{(1-\frac{1}{\alpha})}n(d_2'+\sigma\sqrt{T-t})[e^{-\frac{1}{2}(T-t)\sigma^2(\alpha-1)^2}-1]}{\sigma\sqrt{T-t}},\end{aligned}$$

where $n(\cdot)$ is the Standard Normal pdf function $n(x)\equiv\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$, and m, d_1', and d_2' are as defined above.

It is interesting to note that because d_2' has the same functional form as the original Black-Scholes d_2, then the term $d_2'+\sigma\sqrt{T-t}$ appearing in the delta has the same functional form as the original Black-Scholes d_1. However, this differs from the power call's d_1' which contains an α term.

How does the power call's delta behave as $S(t)$ gets large? Well, as $S(t)$ gets large, both $d_1', d_2' \to \infty$. Thus, $N(d_1')\to 1$, and $n(d_2'+\sigma\sqrt{T-t})\to 0$. It follows that

$$\Delta_{power\ call}\approx \alpha S^{(\alpha-1)}e^{m(T-t)},\quad \text{for large } S(t).$$

It follows that if $S(t)$ is large, then as $(T-t)\to 0$, we get that

$$\Delta_{power\ call}\approx \alpha S^{(\alpha-1)}.$$

This should come as no surprise: If the power call is deep in-the-money, and there is little time to maturity, then its sensitivity to changes in $S(t)$ will be about the same as the sensitivity of $S^\alpha(t)$ to changes in $S(t)$. The latter sensitivity is just

$$\frac{\partial S^\alpha(t)}{\partial S(t)}=\alpha S^{(\alpha-1)}.$$

One implication of this is that the delta of a power call continues to change as $S(t)$ increases.

What may come as a surprise is the shape of the power call option pricing function (see figure B.5). If $\alpha>1$, the plot of $c(t)$ versus $S(t)$ is *steeper* than and above the plot of $\max(S^\alpha-X,0)$ for large $S(t)$ (it decays down toward the payoff as maturity approaches). If $\alpha<1$, the plot of $c(t)$ versus $S(t)$ is less steep than and *below* the plot of $\max(S^\alpha-X,0)$ for large $S(t)$ (it decays up toward the payoff as maturity approaches). Only in the case $\alpha=1$ do the results agree with those for the standard call: The plot of call value as a function of stock price is less steep than and above the plot of $\max(S-X,0)$. In all cases, the plot of $c(t)$ as a function of $S(t)$ is above the plot of $\max(S^\alpha-X,0)$ for small $S(t)$.

Mathematically, the approximation $\Delta_{power\ call}\approx\alpha S^{(\alpha-1)}e^{m(T-t)}$ drives the results for large $S(t)$ (together with the fact that m is positive if $\alpha>1$ and

negative if $\alpha<1$). Economically, the time value of the option drives the results. When $\alpha>1$, the power of S is so high that the option value grows more quickly with increasing S than does the intrinsic value. When $\alpha<1$, the option value grows less quickly than does the intrinsic value, and the European nature of the option means that there is "negative time value" for having to wait for such a low payout.

The payoff diagram for the power call is a little strange because the "kink" does not occur at $S=X$, but at $S=X^{\frac{1}{\alpha}}$—see figure B.5. For example, if $\alpha=2$, the payoff diagram is flat until $S(T)=\sqrt{X}$ and then is an upward sloping portion of the parabola $S^2(T)-X$. If $\alpha>1$, the delta of the power call will be higher than the delta of a standard call with strike $X^{\frac{1}{\alpha}}$ because the payoff diagram is steeper. Conversely, if $\alpha<1$, the delta of the power call will be lower than the delta of a standard call with strike $X^{\frac{1}{\alpha}}$ because the payoff diagram is less steep.

In the power option pricing formulae, d_2' has the same functional form as the d_2 in the regular Black-Scholes. The only difference is that you have $\ln\left(\frac{S(t)}{K}\right)$, where $K=X^{\frac{1}{\alpha}}$, in place of $\ln\left(\frac{S(t)}{X}\right)$. The reasoning follows a Z-score argument (see details in Crack [2009]). In the standard Black-Scholes formula, $N(d_2)$ is the (risk-neutral) probability that the call finishes in-the-money; it is the probability that $S(T)>X$. In the power call option formula, $N(d_2')$ is the (risk-neutral) probability that the power call finishes in-the-money. For the power call, this is the probability that $S^\alpha(T)>X$. This is the same as the probability that $S>X^{\frac{1}{\alpha}}$. This probability is in turn just the standard $N(d_2)$, in the case where the strike is given by $K\equiv X^{\frac{1}{\alpha}}$.

To extend the formulae to the case of continuous dividends at rate ρ, replace $S(t)$ by $S(t)e^{-\rho(T-t)}$ throughout the power option pricing formulae to yield

$$\begin{aligned}c(t) &= S^\alpha(t)e^{(m-\alpha\rho)(T-t)}N(d_1')-e^{-r(T-t)}XN(d_2'), \text{ and}\\ p(t) &= e^{-r(T-t)}XN(-d_2')-S^\alpha(t)e^{(m-\alpha\rho)(T-t)}N(-d_1'), \quad \text{where}\\ d_1' &= \frac{\ln\left(\frac{S(t)}{K}\right)+(r-\rho+(\alpha-\frac{1}{2})\sigma^2)(T-t)}{\sigma\sqrt{T-t}},\\ d_2' &= \frac{\ln\left(\frac{S(t)}{K}\right)+(r-\rho-\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}=d_1'-\alpha\sigma\sqrt{T-t},\\ K &\equiv X^{\frac{1}{\alpha}}, \quad\text{and}\quad m\equiv\left(r+\frac{\alpha}{2}\sigma^2\right)(\alpha-1).\end{aligned}$$

Story: Within two minutes I could see that a guy I was interviewing had inflated his CV. Your CV is in the cross-hairs. Putting something false on it is a stupid thing to do! Do not do it!

SECOND SOLUTION
An alternative to the full and formal pricing formulae given is an approximation using the standard Black-Scholes formula. Simply use d_2' exactly as above (with $K = X^{\frac{1}{\alpha}}$ for the reasons given), use $d_1' = d_2' + \sigma\sqrt{T-t}$, and replace $S(t)$ by $S^\alpha(t)$ in each of $c(t)$ and $p(t)$. However, be warned, this is an approximation only. If α is far from one (say above 1.2 or below 0.8), or time to maturity is longer than about six months, or implied volatility is bigger than about 0.40, the approximation is poor.

JARROW AND TURNBULL'S POWERED CALL
Jarrow and Turnbull ask their readers to value a call with payoff $[S(T)-K]^2$ if $S(T) \geq K$ and zero otherwise (Jarrow and Turnbull [1996, p175]). Assuming a Black-Scholes world, it is easy to show that the value of this call at time t prior to maturity is

$$\begin{aligned} c(t) &= S^2(t)e^{(r+\sigma^2)(T-t)}N(d_0) - 2KS(t)N(d_1) \\ &\qquad + e^{-r(T-t)}K^2N(d_2), \quad \text{where} \\ d_l &= \frac{\ln\left(\frac{S(t)}{K}\right) + \left(r + \left[\frac{3}{2} - l\right]\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}, \quad \text{for } l = 0, 1, 2. \end{aligned}$$

More generally, the following result (Crack [1997])

$$E^*[S^\alpha(T)|S(T) \geq K] = S^\alpha(t)e^{\alpha\left(r+\left[\frac{\alpha-1}{2}\right]\sigma^2\right)(T-t)}N(d_{2-\alpha}),$$

where E^* is expectation with respect to the risk-neutral probability measure and d_l is as above, allows you to value the powered call with general payoff

$$c(T) = \begin{cases} [S(T)-K]^\alpha & ; S(T) \geq K \\ 0 & ; S(T) < K, \end{cases}$$

for non-negative integer α (Crack [1997]). The general pricing formula is

$$\begin{aligned} c(t) &= \sum_{j=0}^{\alpha}(-K)^{\alpha-j}\binom{\alpha}{j}S^j(t)e^{\left[(j-1)\left(r+j\frac{\sigma^2}{2}\right)(T-t)\right]}N(d_{2-j}), \quad \text{where} \\ d_l &= \frac{\ln\left(\frac{S(t)}{K}\right) + \left(r + \left[\frac{3}{2} - l\right]\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}, \quad l = 2, 1, \ldots, 2-\alpha, \end{aligned}$$

and $\binom{\alpha}{j} \equiv \frac{\alpha!}{j!(\alpha-j)!}$ is the usual binomial coefficient (see also Haug [2007, p119]). The reader should check that in the special case $\alpha = 2$, the general formula reduces to that previously given, and that in the special case $\alpha = 1$, the general formula reduces to standard Black-Scholes.[48]

[48] Haug (2007, pp118–119) gives some of these formulae and cites Crack (1997, 2004).

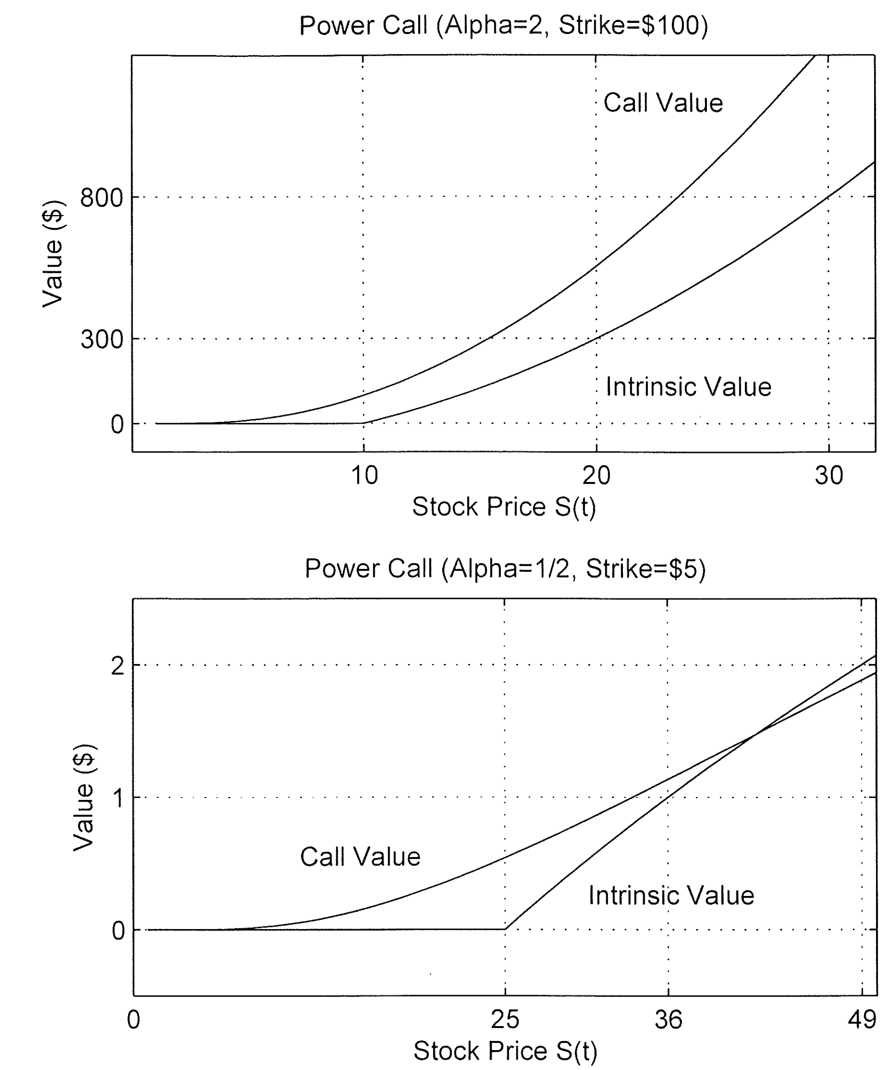


Figure B.5: Power Calls with $\alpha > 1$, and $\alpha < 1$

Note: The power call prices are plotted as a function of price of underlying. Note that the "kink" in the payoff diagram does not occur at the strike K, but rather at $K^{\frac{1}{\alpha}}$ (see Question 2.32).

Answer 2.33: If the Black-Scholes assumptions are correct, then the implied volatilities of options (those backed out of the Black-Scholes pricing formula given the other pricing parameters) should fall on a horizontal line when plotted against strike prices of the options used. However, the patterns that result include smiles and skewed lines depending upon the underlying asset and the time period (Hammer [1989]; Sullivan [1993]; Murphy [1994]; Derman and Kani [1994]). Before the Crash of 1987, you typically got smiles when you plotted the implied volatilities against strikes. Nowadays you are more likely to get skews, or smirks.[49]

What is happening may be viewed in some different and related ways. Option prices are determined by supply and demand, not by theoretical formulae. The traders who are determining the option prices are implicitly modifying the Black-Scholes assumptions to account for volatility that changes both with time and with stock price level. This is contrary to the Black and Scholes (1973) assumption of constant volatility irrespective of stock price or time to maturity. That is, traders assume $\sigma = \sigma(S(t), t)$, whereas Black and Scholes assume σ is just a constant.[50]

If volatility is changing with both level of the underlying and time to maturity, then the distribution of future stock price is no longer Lognormal. The distribution must be something different. Black-Scholes option pricing takes discounted expected payoffs relative to a Lognormal distribution. As volatility changes through time, you are likely to get periods of little activity and periods of intense activity. These periods produce peakedness and fat tails respectively (together called "leptokurtosis"), in stock returns distributions. Fat tails are likely to lead to some sort of smile effect, because they increase the chance of payoffs away-from-the-money.[51]

These irregularities have led to "stochastic volatility" models that account for volatility changing as a function of both time and stock price level (Hull and White [1987]; Scott[1987]; Wiggins [1987]; Hull [1997]). Applications to FOREX options include Chesney and Scott (1989) and Melino and Turnbull (1990). The effect of stochastic volatility on options values is similar to the effect of a jump component: both increase the probability that out-of-the-money options will finish in-the-money and increase the probability that in-the-money options will finish out-of-the-money (Wiggins [1987, pp360–361]). Whether the smile is skewed left, skewed right, or symmetric in a stochastic volatility model depends upon the sign of the correlation between changes in volatility and changes in stock price (Hull [1997, section 19.3]).

[49]Another related deviation from Black-Scholes pricing is that implied volatilities when plotted against term to maturity produce a "term-structure of volatility." That is, traders use different volatilities to value long-maturity and short-maturity options (Derman and Kani [1994, pp2–3]; Hull [1997, pp503–504]).

[50]Black (1976) is the earliest paper I know of that acknowledges that $\sigma \uparrow$ as $S \downarrow$, and vice versa.

[51]The interaction of skewness and kurtosis of returns gives rise to many different possible smile effects (Hull [1997, section 19.3]; Krause [1998, pp145-148]).

Answer 2.34: This question is probably supposed to invoke misleading memories of the barrier option parity relationship: Other things being equal, a down-and-out call plus a down-and-in call is the same as a standard call. However, a double-barrier knock-out is not the same as an up-and-out together with a down-and-out. The latter pair of options is more valuable than the double-barrier knock-out. The most obvious reason is that if the underlying moves one way, then the double knock-out is knocked out, but a portfolio of a down-and-out plus a down-and-in still contains one option. That is, the pair of knock-outs is more versatile—and thus more valuable.

A double-barrier knock-out can be priced using a lattice (e.g., binomial) method. It may also be priced using the Kunitomo-Ikeda formula (Kunitomo and Ikeda [1992]; Musiela and Rutkowski [1997, p211]; or the user-friendly Haug [2007, p156]). The Kunitomo-Ikeda formula is an infinite series. Typically, only the leading few terms are needed for practical purposes (Kunitomo and Ikeda [1992, p286]). More terms may be needed if volatility is high, term to maturity is long, or the distance between the barriers is small (in each case this increases the likelihood of knockout and the pricing is more difficult).

Answer 2.35: The prices of the digital asset-or-nothing, $da(t)$, and the digital cash-or-nothing (with a "bet" size of \$1), $dc(t, \$1)$, are just the two parts of the Black-Scholes formula

$$\begin{aligned} c(t) &= da(t) - X dc(t, \$1), \text{ where} \\ da(t) &= S(t)N(d_1), \\ dc(t, \$1) &= e^{-r(T-t)}N(d_2), \\ d_1 &= \frac{\ln\left(\frac{S(t)}{X}\right) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}, \text{ and,} \\ d_2 &= d_1 - \sigma\sqrt{T-t}. \end{aligned}$$

A Black-Scholes derivation using discounted expected payoffs under risk-neutral probabilities (Crack [2009]) contains implicit derivations of both digital option values. These may be identified if the initial step is re-expressed as

$$\begin{aligned} da(t) &= e^{-r(T-t)}\, E^*[S(T)\mathcal{I}_{S(T)>X} \mid S(t)], \text{ and,} \\ dc(t, \$1) &= e^{-r(T-t)}\, E^*[\mathcal{I}_{S(T)>X} \mid S(t)], \end{aligned}$$

where $\mathcal{I}_{S(T)>X}$ is the indicator function

$$\mathcal{I}_{S(T)>X} = \begin{cases} 1 \text{ if } S(T) > X, \\ 0 \text{ if } S(T) \leq X. \end{cases}$$

Answer 2.36: A path-dependent option is one where the final payoff depends upon the stock price path followed. If the stock price ends up between the barriers, the option has different values, depending upon whether it was knocked in or knocked out (or both). Path-dependent options can typically be priced using Monte-Carlo methods. However, Monte-Carlo does not work for American-style options. Standard lattice techniques (e.g., binomial option pricing) do not usually work for path-dependent options.[52] However, you can price the "out-in" derivative using standard lattice methods, as follows. The parity relationship for knock-outs says that a down-and-out plus a down-and-in is a standard option. We can generalize this to conclude that an out-in plus a double-barrier knock-out is the same as an up-and-out (other things being equal). It follows that the out-in is worth the excess of the value of the up-and-out over the double-barrier knock-out. Both these knock-outs can be priced using standard lattice techniques.

Answer 2.37: Let $G(\cdot)$ denote the gold price. Now is time t, and time T is six months from now. The naive (and incorrect) step is to conclude that a volatility of σ = \$60 per annum translates to a six-month volatility of \$30. In fact, volatility grows with the square root of the term. Thus, \$60 per year translates to about $\sqrt{\frac{1}{2}} \times \$60 \approx \$42$ per half-year.

How do we find the probability that the option finishes in-the-money, $P(G(T) > 430)$? With $r = 0$ there is no drift in the risk-neutral world, so the distribution of $G(T)$ is centered on $G(t) = \$400$, with standard deviation roughly \$42. Thus

$$\begin{aligned} P(G(T) > 430) &= P(G(T) - G(t) > 30) \\ &= P\left(\frac{G(T) - G(t)}{42} > \frac{30}{42}\right) \\ &\approx 1 - N\left(\frac{3}{4}\right), \end{aligned}$$

where $N(\cdot)$ is the cumulative Standard Normal function. The last step follows because $\frac{G(T)-G(t)}{42}$ is roughly Standard Normal. We know that $N(0) = 0.50$, and $N(1) = 0.84$, so $N\left(\frac{3}{4}\right) \approx 0.75$.

We conclude that there is roughly a 25% chance that the digital option finishes in-the-money. With a bet size of \$1 million, and a riskless interest rate of zero, the discounted expected payoff (in a risk-neutral world) is roughly \$250,000. The erroneous σ = \$30 gives an incorrect value of only about \$160,000.

Answer 2.38: Let us review quickly standard American options before looking at the perpetual option. American options are harder to price than European ones. Puts are harder to price than calls. An American put is hardest of all to price because early exercise can in general be optimal at any time for an American-style put. This differs from an American-style call, for which early exercise is optimal only at a few dates during the option's life (just prior to ex-dividend days). In fact, the problem is so hard that no exact pricing formula exists for standard American put options.

[52] Standard lattice techniques can be modified to allow pricing of path-dependent options. However, a couple of conditions involving complexity of the payoffs need to be satisfied (Hull and White [1993]; Hull [1997]).

Black and Scholes (1973) value European-style puts and calls. If a stock does not pay dividends, then a European call and an American call have the same value (there is no incentive to exercise early). Thus, American calls on non-dividend-paying stocks can be valued using Black-Scholes. The introduction of dividends complicates matters. However, an approximate pricing formula (Black [1975]) and an exact pricing formula (Roll [1977b]; Geske [1979]; Whaley [1981]) for American calls on dividend-paying stocks are known (see Hull [1997, chapter 11]). American puts are more complicated. The dividend issue is not as important for puts as for calls because it is the receipt of the strike, not the dividends, that encourages early exercise of a put. Although no exact American put pricing formula exists, there are approximations (Parkinson [1977]; MacMillan [1986]; Barone-Adesi and Whaley [1987]). See the summaries in tables B.2 and B.3.

Now to the perpetual American put. Extending the life of an option in perpetuity eases the pricing burden (removing the dependence on time turns a PDE into an ODE). Pricing the perpetual American put was a question on a problem set I had as a student in Robert C. Merton's derivatives course at Harvard in 1991. I reproduce my solution here.

Let V denote the value of a perpetual American put option on a stock. Let S denote the stock price. Let X denote the strike price. Assume that the stock pays continuous dividends at rate ρ. Let σ and r denote the volatility of stock returns and the riskless interest rate respectively. The Black-Scholes PDE is given by (Wilmott et al. [1993])

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - \rho) S \frac{\partial V}{\partial S} - rV = 0.$$

However, for a *perpetual* put, time decay must be zero (it cannot age if it can live forever). Thus, the PDE becomes an ODE:

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - \rho) S \frac{\partial V}{\partial S} - rV = 0$$

Let $\underline{S}$ denote the lower exercise boundary (this is how low the stock has to go before exercise of the put becomes optimal—it has to be determined). Then, we have the boundary conditions

$$\begin{aligned} V(S = \underline{S}) &= X - \underline{S}, \\ \left.\frac{\partial V}{\partial S}\right|_{S=\underline{S}} &= -1, \\ V(S) &\leq X. \end{aligned}$$

The second condition is the "high contact" condition.

All of this ODE's solutions may be represented as a linear combination of any two linearly independent solutions. It follows that

$$V(S) = A_1 V^1(S) + A_2 V^2(S),$$

where A_1 and A_2 are constants, and V^1 and V^2 are linearly independent solutions of the ODE. My guess is that $V^1 = S^{\lambda_1}$, and $V^2 = S^{\lambda_2}$ for some constants λ_1, and λ_2.[53] Substitution of V^i into the ODE yields (for $i = 1,\ 2$, and for $\underline{S} \leq S$)

$$\left[\frac{1}{2}\sigma^2\lambda_i(\lambda_i - 1) + (r-\rho)\lambda_i - r\right] S^{\lambda_i} = 0.$$

Rearranging and collecting terms in λ_i, we get for $i = 1,\ 2$

$$\frac{1}{2}\sigma^2\lambda_i^2 + \left(r - \rho - \frac{1}{2}\sigma^2\right)\lambda_i - r = 0.$$

This is a quadratic formula, with solutions for λ_i:

$$\begin{aligned} \lambda_1 &= \frac{-\left(r-\rho-\frac{1}{2}\sigma^2\right) + \sqrt{\left(r-\rho-\frac{1}{2}\sigma^2\right)^2 + 2\sigma^2 r}}{\sigma^2}, \quad \text{and} \\ \lambda_2 &= \frac{-\left(r-\rho-\frac{1}{2}\sigma^2\right) - \sqrt{\left(r-\rho-\frac{1}{2}\sigma^2\right)^2 + 2\sigma^2 r}}{\sigma^2} \end{aligned}$$

The solutions for λ_i can be seen to satisfy $\lambda_1 > 0$ if $r > 0$, and $\lambda_2 < 0$ if $r > 0$. Let us now consider the behaviour of the general solution we have derived: $V(S) = A_1 S^{\lambda_1} + A_2 S^{\lambda_2}$. First of all, with $\lambda_1 > 0$, and $\lambda_2 < 0$, then

$$\lim_{S\to+\infty}\left(A_1 S^{\lambda_1} + A_2 S^{\lambda_2}\right) = \pm\infty, \quad \text{if } |A_1| > 0.$$

However, the boundary conditions put both upper and lower finite bounds on the value of the put. Therefore, $A_1 = 0$, and $V(S) = A_2 S^{\lambda_2}$. Now, the first boundary condition tells us that

$$V(\underline{S}) = A_2 \underline{S}^{\lambda_2} = X - \underline{S},$$

so it follows that $A_2 = \frac{X-\underline{S}}{\underline{S}^{\lambda_2}}$, which yields

$$V(S) = \left(\frac{X-\underline{S}}{\underline{S}^{\lambda_2}}\right) S^{\lambda_2} = (X - \underline{S})\left(\frac{S}{\underline{S}}\right)^{\lambda_2}.$$

[53]Why make this guess? Look at the ODE: the degree of the derivatives of V and the degree of S in the coefficients move together (both two, then both one, then both zero). This suggests solutions that are powers of S.

To pinpoint the solution, we must determine the value of the lower exercise boundary $\underline{S}$. The second of our boundary conditions says $\frac{\partial V}{\partial S}\big|_{S=\underline{S}} = -1$. We can solve for $\underline{S}$ using this.

$$\begin{aligned} \frac{\partial V}{\partial S} &= \lambda_2(X-\underline{S})\left(\frac{S^{\lambda_2-1}}{\underline{S}^{\lambda_2}}\right) \\ \Rightarrow \quad \frac{\partial V}{\partial S}\bigg|_{S=\underline{S}} &= \lambda_2\frac{(X-\underline{S})}{\underline{S}} = -1 \\ \Rightarrow \lambda_2(X-\underline{S}) &= -\underline{S} \\ \Rightarrow \quad \underline{S} &= \frac{\lambda_2 X}{\lambda_2 - 1}. \end{aligned}$$

Thus, for $S \geq \underline{S} \equiv \frac{\lambda_2 X}{\lambda_2 - 1}$, the perpetual American put is worth

$$\begin{aligned} V(S) &= (X-\underline{S})\left(\frac{S}{\underline{S}}\right)^{\lambda_2} = \left(\frac{X}{1-\lambda_2}\right)\left[\frac{(\lambda_2-1)S}{\lambda_2 X}\right]^{\lambda_2}, \quad \text{where} \\ \lambda_2 &= \frac{-\left(r-\rho-\frac{1}{2}\sigma^2\right) - \sqrt{\left(r-\rho-\frac{1}{2}\sigma^2\right)^2 + 2\sigma^2 r}}{\sigma^2}. \end{aligned}$$

For $0 \leq S \leq \frac{\lambda_1 X}{(\lambda_1 - 1)} \equiv \overline{S}$, it may be shown using similar techniques that a perpetual American call is worth

$$\begin{aligned} V(S) &= \left(\frac{X}{\lambda_1 - 1}\right)\left[\frac{(\lambda_1-1)S}{\lambda_1 X}\right]^{\lambda_1}, \quad \text{where} \\ \lambda_1 &= \frac{-\left(r-\rho-\frac{1}{2}\sigma^2\right) + \sqrt{\left(r-\rho-\frac{1}{2}\sigma^2\right)^2 + 2\sigma^2 r}}{\sigma^2}. \end{aligned}$$

Note that (theoretically at least) a perpetual European call is worth the same as the stock, whereas a perpetual European put is worth zero (look at the limiting behaviour of the Black-Scholes formula).[54]

Answer 2.39: If you subtract LIBOR, denoted "L," from both payments, it seems that Party B is paying $24\% - 3 \times L$. This is three times $8\% - L$. The quoted swap is, therefore, equivalent to three swaps, each of which is a swap of LIBOR for 8% fixed (where Party A pays LIBOR, and Party B pays 8%).

Answer 2.40: If you sold the option, you should hold about one-half a share to hedge. If you bought the option, you should short about one-half a share to hedge. If you are at-the-money, there is about a fifty-fifty chance the option finishes in-the-money; and with this expectation, you need about one-half a share to hedge.

[54]Note that in the case of the perpetual American call, $\lim_{\rho\to 0}\lambda_1 = 1$, and $\lim_{\lambda_1\to 1} V(S) = S$. That is, with no dividends, the perpetual American call has the same value as the stock—just like the perpetual European call.

Table B.2: Pricing Methods Summary: Plain Vanilla Options

	European-Style		American-Style	
	Put	Call	Put	Call
No dividends	Black-Scholes put formula	Black-Scholes call formula	No exact formula (use approximation formula, tree, or finite differences)	Black-Scholes call formula (early exercise is never optimal)
Lump sum dividend D	Use $S^*=S-PV(D)$ in Black-Scholes	Use $S^*=S-PV(D)$ in Black-Scholes	No exact formula (use approximation formula, tree, or finite differences)	Roll-Geske-Whaley formula, or Black's pseudo formula
Continuous dividends at rate ρ	Use $S^*=Se^{-\rho(T-t)}$ in Black-Scholes (Merton's formula)	Use $S^*=Se^{-\rho(T-t)}$ in Black-Scholes (Merton's formula)	No exact formula (use approximation formula, tree, or finite differences)	Adjust Roll-Geske-Whaley formula
$S=(\frac{USD}{FX})$	Use $\rho = r_{FX}$ in Merton's formula (Garman-Kohlhagen/Grabbe formula)	Use $\rho = r_{FX}$ in Merton's formula (Garman-Kohlhagen/Grabbe formula)	Use $\rho = r_{FX}$ in the above	Use $\rho = r_{FX}$ in the above
All cases: Numerical	Monte Carlo, lattice, or finite differences		Lattice or finite differences	

Note: Pricing methods for European- or American-style plain vanilla puts or calls where the underlying pays no dividends, pays a lump sum dividend, pays continuous dividends, or is a foreign currency.

Table B.3: Pricing Methods Summary: Exotic Options

European-Style		American-Style	
Path-Independent	Path-Dependent	Path-Independent	Path-Dependent
Lattice, Monte Carlo, or finite difference	Monte Carlo, finite difference, lattice (difficult)	Lattice or finite differences	Lattice (difficult) or finite differences
... or a formula if you can derive it			

Note: Summary of pricing methods for exotic options that are European- or American-style, path-independent or path-dependent.

Answer 2.41: Mean reversion is the tendency for a variable to return to some sort of long-run mean. Interest rates are generally considered to be mean-reverting: they go up, they go down, but they eventually return to some sort of long-term average. In the case of a mean-reverting stock price, the stock price would tend to be pulled back to the average if the price rises or falls very far. This may reduce volatility and make the option cheaper.

A model of mean reversion makes sense for interest rates, and for stock returns,

but it is by no means clear to me that it makes sense for stock *prices*. Bates argues that strong mean reversion in stock prices is implausible because of speculative opportunities available from buying when $S < \bar{S}$ and selling when $S > \bar{S}$ (Bates [1995, pp7–8]). Lo and Wang say that autocorrelation in asset *returns* can increase or decrease σ (and the option price) and that it depends upon the specification of the drift in the model (Lo and Wang [1995, p105]).

Mean reversion is really just negative autocorrelation at some horizon. At short horizons (e.g., daily or weekly), stock index returns are positively autocorrelated (Lo and MacKinlay [1988]). At longer horizons (e.g., three or four years), Fama and French (1988) and Poterba and Summers (1988) say that stock returns are negatively autocorrelated (i.e., mean reverting). However, evidence for this is weak (Richardson [1993]). Lo and MacKinlay (1988, p61) say that longer-term positive autocorrelation is not inconsistent with shorter-term negative autocorrelation (i.e., mean reversion). Peterson et al. (1992) and Lo and Wang (1995) discuss option pricing when asset returns are autocorrelated. Crack and Ledoit (2008) discuss hypothesis testing when asset returns are autocorrelated.

Answer 2.42: Hedging can increase your risk if you are forced to both buy short-dated options and hedge them. In this case, to hedge, you need to short the stock. If the stock price rises up to the strike, and the options (be they puts or calls) expire worthless, then you lose on both the options and the short stock position. By hedging, you end up worse off than if you had not hedged.

Answer 2.43: This is a common question. You can hedge the written put by shorting an asset whose returns are correlated with returns on the underlying stock. Ideally, this would just be the stock itself. However, it is not always possible to short stock. Shorting some index futures would give you an (imperfect) hedge. You need to know either the beta or the correlation of the stock relative to the index to apportion the hedge correctly. You could also short the stock of a close competitor.

Answer 2.44: People are fed by the area, A, of the pizza. $A = \pi r^2 = \pi\left(\frac{d}{2}\right)^2 = \frac{\pi}{4}d^2$, where d is the diameter. Thus, $d = \sqrt{\frac{4}{\pi}}\sqrt{A}$. Multiplying A by $\frac{8}{6}$ requires a multiplicative change of $\sqrt{\frac{8}{6}}$ in d. That is, $d' = \sqrt{\frac{8}{6}}d = 13.86$ inches. Without a calculator, the square root of $(1+X)$ is roughly $(1+\frac{X}{2})$, so $\sqrt{\frac{8}{6}} \approx \sqrt{1.33} \approx 1.15$. Fifteen percent of 12 is 1.8, so the answer is roughly 13.8 inches.

Why is this a derivatives question? Using the approximation $c = S\sigma\sqrt{\frac{T-t}{2\pi}}$, a question with the same answer is: a six-month at-the-money call has price \$12; what is the price of the eight-month call?

Answer 2.45: You want to be short a put if you expect the put to fall in price (e.g., the underlying is expected to rise, volatility to expected to fall, etc).

Answer 2.46: A fair price for future delivery of an asset depends upon the spot price and the cost of carry. The cost of carry includes the cost of money (i.e., an interest rate), dividend income, storage costs, and the convenience yield. The only difference between the two pieces of land is the entrance fee to the beach. This is a dividend that lowers the forward price of the beach relative to the field.

Answer 2.47: There are two important points: use of logarithms, and division by $T-1$. Begin by calculating continuously compounded returns (as used in Black-Scholes):

$$\begin{aligned} X_t &\equiv \ln(1+R_t) \\ &= \ln\left(1+\frac{P_t - P_{t-1}}{P_{t-1}}\right) \\ &= \ln\left(\frac{P_t}{P_{t-1}}\right) \end{aligned}$$

With 30 stock prices, you get $T = 29$ returns. Now calculate the standard sample mean and variance. Remember to divide by $T-1 = 28$ in the variance estimator to get an unbiased small sample estimator of historical volatility (DeGroot [1989, p413]).

$$\begin{aligned} \hat{\mu} &= \frac{1}{T}\sum_{t=1}^{T} X_t \\ \hat{\sigma}^2 &= \frac{1}{T-1}\sum_{t=1}^{T}(X_t - \hat{\mu})^2. \end{aligned}$$

Some people may even leave off the "$-\hat{\mu}$" in the $\hat{\sigma}^2$ calculation because mean daily stock returns are typically so tiny compared to volatility, but I prefer to leave it in.

Answer 2.48: The key is default risk, but let's start with a quick swap curve review. Swap rates are fixed rates quoted by dealers against the floating leg (e.g., six-month USD LIBOR) of an interest rate swap. The "swap buyer" is the fixed-rate payer and is said to be "long the swap" (although I have also heard the reverse). The swap curve is inferred from quoted swap rates for different maturities in the same manner that a zero-coupon yield curve (i.e., a "spot curve") is bootstrapped from the yields on coupon bearing bonds of different maturities. Swaps dealers can do customized deals offering different quoted swap rates to companies of different credit rating; however, dealers tend to quote the same swap rate to companies of different credit rating but ask for different amounts of collateral based on the rating (personal communication with a NY dealer [April, 1999]).[55] The collateral and subsequent margin calls

[55] Minton (1997, p252) confirms that the plain vanilla swap quotes in her (1992 and earlier) sample assume no credit enhancement (e.g., margins or marking to market).

essentially resolve the credit issues. Johannes and Sundaresan (2003, p9) state that the "key to effective credit risk mitigation is frequent margin calls." They state that more than 65% of plain vanilla derivatives, and especially interest rate swaps, are collateralized and that at least 74% mark to market at at least a daily frequency.

The settlement features of an interest rate swap mean that default risk in a swap is higher than in a eurodollar futures contract but lower than in a bond (Minton [1997, p253]). The reasoning is as follows. The settlement rate for the futures contract is reset daily by market forces, but the swap typically resets only every six months. Both the futures contract and the swap are marked-to-market and use margins, but the futures contract is backed by the triple-A-rated futures clearing corporation as a counterparty of last resort and so the futures contract is less credit-risky than the swap. The swap differs from the bond because no principal changes hands.[56] At initiation, the value of a swap contract is zero; but during the life of the swap, as interest rates rise and fall, the value of the contract can become positive or negative, respectively, to the swap buyer. Although a bondholder is always worried about default risk, the swap buyer worries about default risk only when the swap has positive value. Default on a swap is thus less likely than default on a bond because default on a bond requires only that the company be in financial distress, whereas default on a swap requires both that the company be in financial distress and that the remaining value of the swap be positive. The joint probability of both events needed for swap default is less than the single probability needed for bond default (Minton [1997, p262–263, p267]).

It follows that the coupon rate on a bond will be higher than the quoted swap rate for a swap of the same maturity. This is true for all maturities, so bootstrapping the swap curve from swap rates of swaps of different maturities and bootstrapping the zero-coupon yield curve (i.e., the spot curve) from coupon rates of bonds of different maturities produces a swap curve strictly below the zero-coupon yield curve.[57] It follows that when you discount the cash flows to the bond using the swap curve, you get a number above that which you would get when you discount the cash flows to the bond using the zero-coupon yield curve (i.e., above par).

Answer 2.49: Try a simple economics argument. The option must cost the same as a replicating portfolio—else there is money to be made. This result is driven by no-arbitrage and is thus independent of risk preferences. I can ease my calculations by assuming risk-neutrality for everyone in the economy. In

[56] Although true for an interest rate swap, this is not so in a forex swap, where principal changes hands at initiation and conclusion of the life of the swap.

[57] In early 1999, just before the Tech Bubble, two-year swap rates were about 40 bps higher than US treasuries (which were at 500 bps), about 5 bps lower than the yields on AAA-rated debt, about 30 bps lower than the yields on AA-rated debt and about 40 bps lower than the yields on BBB-rated debt. For comparison, ten years later in mid-2009, coming out of the Global Credit Crisis, these rates/spreads were 45 bps, 120 bps, 3 bps, 75 bps, and 575 bps respectively.

such an economy in equilibrium, the required return (and thus the expected growth rate and also the discount rate) for all traded securities is the riskless rate. I price the option as if we are in this economy (and the option pricing is immune to this assumption).

Answer 2.50: Options live in the future, not the past: Today is the first day of the rest of the life of a traded option. Setting aside problems with volatility smiles and skews, the implied volatility (or "implied standard deviation") is a market-consensus forecast of volatility over the remaining life of the option. It would be logical, therefore, that implied volatility is a better predictor of future volatility than is historical volatility. Indeed, this is found empirically for both FOREX (Xu and Taylor [1995]) and for equity indices (Fleming [1998]).

Answer 2.51: Assume that $S = \$100$ so that the call has strike $X_c = 110$, and the put has strike $X_p = 90$. The distribution of future stock prices is Lognormal in the Black-Scholes model and is thus skewed with its mean higher than the median, which is in turn higher than the mode (the peak). The median of the distribution of future stock prices is $Se^{(r-\frac{1}{2}\sigma^2)}$. The term $(r - \frac{1}{2}\sigma^2)$ tends to be close to zero, so the median is approximately S. It follows that $X_c = 110$ and $X_p = 90$ are roughly equidistant from the median. Half the distribution is above the median and half is below. The probability that the call finishes in-the-money is $P(S(T) > X_c) = P(S(T) > 110) \approx \frac{1}{2} - P(100 < S(T) < 110)$. The probability that the put finishes in-the-money is $P(S(T) < X_p) = P(S(T) < 90) \approx \frac{1}{2} - P(90 < S(T) < 100)$. However, the distribution is so skewed that $P(90 < S(T) < 100) \gg P(100 < S(T) < 110)$ (at the median the probability density function is downward sloping). Thus, $P(S(T) > X_c) \gg P(S(T) < X_p)$, and the call is more likely to finish in-the-money than the put, and this typically makes the call more valuable.

Answer 2.52: The value of the derivative, V, must satisfy the Black-Scholes PDE (Wilmott et al. [1993]):

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

These derivatives are just the theta (Θ), gamma (Γ), and delta (Δ), respectively, so we rewrite the PDE as

$$\Theta + \frac{1}{2}\sigma^2 S^2 \Gamma + rS\Delta - rV = 0.$$

The last two terms may be written as $r(S\Delta - V)$, and they offset to some extent. The entire PDE adds to zero, so that leaves $\Theta + \frac{1}{2}\sigma^2 S^2 \Gamma$ taking a value close to zero. This means that Θ and Γ are typically going to be of opposite signs. Not only that, but their magnitudes are going to be correlated. For example, if Θ is large and negative then Γ is probably large and positive (e.g., an at-the-money call close to maturity has these properties).

There are two exceptions amongst plain vanilla puts and calls. These were mentioned on page 106: Θ can be positive for a deep in-the-money European call (if the dividend yield is high enough), and Θ can also be positive for a deep in-the-money European put. As long as these options are not so deep in-the-money that they have zero Γ, then they can have both positive Θ and positive Γ.

Answer 2.53: If you said you have an 80% chance of getting \$20, and a 20% chance of getting nothing, giving an expected payoff of \$16, which you then discount at zero to get an answer of \$16 for the call value, you are wrong! Sure enough, the call does have an expected payoff of \$16 in the real world, but the discount rate is not zero. The discount rate is some leveraged version of the discount rate on the stock, and you do not have that information. Try again, then come back here for the answer below.

We do not know the discount rate on the stock. We do not know the discount rate on the option. We must use risk-neutral valuation. The risk-neutral probability π^* of an up move in the stock satisfies

$$\begin{aligned} S &= e^{-r(\Delta t)}\left[\pi^* Su + (1-\pi^*)Sd\right], \\ \text{that is, } \$100 &= \pi^*\$130 + (1-\pi^*)\$70, \end{aligned}$$

where r is the riskless rate (zero here), u is the multiplicative "up" growth factor in the stock (1.30 here), and d is the multiplicative "down" growth factor in the stock (0.70 here). See Crack (2009) or you favorite option pricing book for deeper details of binomial/lattice pricing. Simple algebra yields $\pi^* = 0.50$. The value of the call is then

$$\begin{aligned} c &= e^{-r(\Delta t)}\left[\pi^* \max(0, Su - X) + (1-\pi^*)\max(0, Sd - X)\right] \\ &= 1 \cdot [0.50 \cdot (\$130 - \$110) + 0.50 \cdot (\$0)] = \$10. \end{aligned}$$

Story: 1. Man wore jogging suit to interview for position as financial vice president. 2. Interrupted to phone his therapist for advice on answering specific interview questions.

Interview Horror Stories from Recruiters
Reprinted by kind permission of *MBA Style Magazine*

Answer 2.54: The product call pricing formula is so simple that you could simply say "here is the answer, it looks like regular dividend-adjusted Black-Scholes but you replace $S(t)$ by the product $S_1(t) \times S_2(t)$, and you replace σ by $\sigma' \equiv \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}$ where ρ is the instantaneous correlation between the Wiener processes (i.e., Brownian motions) driving S_1 and S_2, and of course the answer is symmetric in S_1, S_2, and their associated 'dividend yields'." However, the derivation is very instructive in risk-neutral pricing, PDE's, and similarity solutions, and I cannot find it in my books so I think it belongs here.

One application of the product call is to the pricing of foreign equity options struck in a domestic currency (Haug [2007, pp226–228]). For example, a US investor has the right to buy one share of NTT corporation stock (trading in Tokyo at JPY price S_2), but the call option strike price is in USD.[58] In this case, the payoff is $\max[S_1(T) \times S_2(T) - X, 0]$, where S_1 is the $\frac{USD}{JPY}$ exchange rate, S_2 is the JPY price of NTT per share, and T is the expiration date.

Make the following definitions:

$$
\begin{aligned}
S_1(t) &= \frac{\text{USD}}{\text{JPY}}(t) \\
S_2(t) &= \frac{\text{JPY}}{\text{Share of NTT}}(t) \\
r_{US} &= \text{US riskless interest rate} \\
r_{JP} &= \text{Japanese riskless interest rate} \\
q &= \text{NTT's continuous dividend yield} \\
dS_1 &= r_1 S_1 dt + \sigma_1 S_1 dw_1 \\
dS_2 &= r_2 S_2 dt + \sigma_2 S_2 dw_2 \\
\sigma_1 &= \text{Volatility of } dS_1/S_1 \text{ process} \\
\sigma_2 &= \text{Volatility of } dS_2/S_2 \text{ process} \\
r_1 &= \text{Drift of } dS_1/S_1 \text{ process} \\
r_2 &= \text{Drift of } dS_2/S_2 \text{ process} \\
\rho dt &= E[(dw_1)\cdot(dw_2)] \;=\; \text{instantaneous correlation} \\
X &= \text{USD-denominated strike price}
\end{aligned}
$$

So, what exactly are r_1 and r_2 in a risk-neutral world? The answer depends upon whether we look from a US or a Japanese perspective (Hull [1997, p301]). We shall use the US perspective. For S_1 from the US perspective, the risk-neutral process has $r_1 = r_{US} - r_{JP}$. For S_2 from the Japanese perspective, $r_2 = r_{JP} - q$, but from the US perspective, $r_2 = r_{JP} - q + (-\rho)\cdot\sigma_1\sigma_2$, where $-\rho$ is the instantaneous correlation between the Wiener processes driving the two JPY-denominated processes $S_2(t)$ and $\frac{JPY}{USD}(t)$. This correlation is the negative of that between the Wiener processes driving $S_2(t)$ and $S_1(t) = \frac{1}{\frac{JPY}{USD}(t)}$ (Hull [1997, p301]). Thus, the risk-neutral drifts from the US perspective are

$$r_1 = r_{US} - r_{JP}, \text{ and } r_2 = r_{JP} - q - \rho\sigma_1\sigma_2,$$

but we shall continue to work with r_1, and r_2, and then plug these in at the end. From our stochastic calculus training we know that as long as dynamic replication is possible, then de-trended prices of traded assets are martingales

[58] Please note that this is *not* a quanto option. Quantos are currency translated options, and so is this, but a quanto takes the JPY price of the foreign security and simply replaces the JPY symbol with a USD symbol when calculating the payoff (Haug [2007, p228]; Hull [1997, p298]; Wilmott [1998, p155]). The JPY security payoff is said to be "quantoed" into USD.

in the risk-neutral economy (Huang [1992]; Crack [2009, section 4.4]). A bullet-point review is called for before proceeding.

• *RISK – NEUTRAL PRICING REVIEW* •

- The technical requirement for dynamic replication to be possible is described nicely in Jarrow and Rudd (1983). Essentially, it requires that for very small time horizons the value of the derivative and the value of the underlying(s) be perfectly linearly correlated. A diffusion or a simple jump process satisfies this, but if the underlying stock price follows a jump-diffusion process (regardless of whether the jump size is deterministic, stochastic, diversifiable, or non-diversifiable), then a replicating portfolio cannot be formed, and the no-arbitrage pricing method fails (Cox and Rubinstein [1985, chapter 7]; Merton [1992]).
- If dynamic replication is possible, then by no-arbitrage the value of the derivative equals the start-up cost of a replicating portfolio.
- If the replication recipe is known (perhaps via an equilibrium CAPM pricing approach as in the original Black and Scholes [1973] paper), then no two economic agents can disagree on the correct arbitrage-free price of the derivative. Thus, regardless of what we assume about the preferences of the agents in the economy, the pricing of the derivative will be the same.
- We ease our calculations substantially by proceeding *as if* the agents in the economy are risk-neutral.[59] That is, although they see the risk, they ignore it completely.
- People in a risk-neutral economy care only about expected return. In equilibrium all traded assets must offer the same expected return (or investors would still be shorting low-yield securities to invest in high-yield ones and we would not yet be in equilibrium). Existence of a government-backed fixed-rate riskless asset means that the riskless rate is the equilibrium required return on all securities in this hypothetical world.
- If risk is not priced by agents in the economy, then traded security prices (including derivatives) are simply discounted expected payoffs where discounting uses the riskless rate, and all traded security prices are assumed to drift upwards at the riskless rate (less any dividend yield, of course—so that total yield is the riskless rate). If risk were priced, then discount rates would need to be risk adjusted, perhaps via the CAPM (Arnold, Crack and Schwartz [2009, 2010]).
- Let $B(t) \equiv e^{rt}$ denote the price of a riskless money market instrument (i.e., you invest \$1 at time 0, and it grows at riskless rate r). Then $B(t)$ drifts upward at the riskless rate. The money market account serves as a benchmark for performance in both the real and risk-neutral worlds.

[59] Important: We are not assuming anyone is really risk-neutral. It is simply that options prices are immune to assumptions about risk preferences, and this proves to be a very helpful assumption.

It seems natural to express other asset prices in terms of units of this asset.[60] That is, instead of looking at security price $P(t)$, look at $\frac{P(t)}{B(t)}$.

- With $B(t)$ drifting upward at the riskless rate, and $P(t)$ expected to drift upward at the same rate in equilibrium in the risk-neutral world, it follows that $\frac{P(t)}{B(t)}$ is expected to have no drift. That is, for any $\Delta t > 0$,

$$E^*\left[\frac{P(t+\Delta t)}{B(t+\Delta t)}\left|\frac{P(t)}{B(t)}\right.\right] = \frac{P(t)}{B(t)},$$

where E^* denotes expectation in the risk-neutral world.

- Let $P^\dagger(t) \equiv \frac{P(t)}{B(t)}$, then the previous result says that for any $\Delta t > 0$,

$$E^*\left[P^\dagger(t+\Delta t)\left|P^\dagger(t)\right.\right] = P^\dagger(t).$$

That is, the best guess of where $P^\dagger$ will be in the future (in the risk-neutral world) is where it is today. This is akin to the efficient markets hypothesis. A random variable with this property is called a "martingale."

- When we assume that traded securities' prices have required returns equal to the riskless rate in the risk-neutral world, we are really just redistributing the probabilities we associate with possible final security price outcomes.[61] However, some things stay the same. For example, if a stock price outcome occurs with probability **0** in the real world, then it still occurs with probability **0** in the risk-neutral world (thus, the range of possible outcomes does not change, only their probability of occurrence; and the transformation of probabilities moves the expected return on IBM, say, from 12% per annum to whatever the T-bill yield happens to be). Similarly, if a stock price outcome occurs with probability **1** in the real world, then it still occurs with probability **1** in the risk-neutral world.
- In probability theory, a mathematical function that allocates probability weight to outcomes in the sample space is called a "measure." Two measures that reassign probabilities to outcomes without changing the range of possible outcomes (as above) are called "equivalent measures."[62]
- Thus, in the risk-neutral world, we reallocate probabilities in an equivalent manner (i.e., same range of possible outcomes), and the price of any traded asset—when "de-trended" by the money market account—follows a martingale. The probability measure (i.e., allocation of probabilities to outcomes) in the risk-neutral world is thus called an "equivalent martingale measure." You see this expression in the more advanced literature.

[60] This is referred to as a change of "numeraire." A numeraire is a base unit of measurement. This is similar to changing units of measurement from USD to GBP, say, except that here we choose a USD-denominated money market account instead of GBP.

[61] Note the word "traded" here. A futures price, for example, is not the price of a traded asset, so its drift need not be r.

[62] The relationship between the two measures is captured by the Radon-Nikodym derivative. See Baxter and Rennie (1998, p65) for simple intuition, Musiela and Rutkowski (1992, pp114, 121) for the advanced mathematics, and Arnold, Crack and Schwartz (2010) for an application.

- Two natural derivative pricing methods fall out of all of the above. The first uses discounted expected payoffs, the second uses PDE's.
- First Method (Cox and Ross [1976]): The martingale property applied to de-trended derivative price V (i.e., $V^\dagger = V/B = Ve^{-rt}$) implies

$$\begin{aligned} V^\dagger(t) &= E^*\left[V^\dagger(T)\left|V^\dagger(t)\right.\right] \\ \Rightarrow V(t)e^{-rt} &= E^*\left[V(T)e^{-rT}\left|V(t)\right.\right] \\ \Rightarrow V(t) &= e^{-r(T-t)}E^*\left[V(T)\left|V(t)\right.\right]. \end{aligned}$$

I derive Black-Scholes in Crack [2009] using precisely this approach: discounted expected payoff in a risk-neutral world.

- Second Method (Harrison and Kreps [1979]): Let V be the derivative price we seek, then the martingale property applied to de-trended V (i.e., $V^\dagger = V/B = Ve^{-rt}$) implies that $dV^\dagger$ has no time trend (i.e., no drift). We can apply Itô's Lemma to $V^\dagger$ to calculate

$$dV^\dagger = [\text{time trend}]dt + \sum_i [\text{diffusion coefficients}]_i dw_i,$$

where dw_i is the i^{th} Brownian motion driving the underlyings. If V is a function of $S(t)$ and t only, and $dS(t) = rSdt + \sigma Sdw$ then

$$\begin{aligned} dV^\dagger(S(t),t) &= d[V(S(t),t)e^{-rt}] \\ &\overset{\text{Itô}}{=} \left(V_S dS + V_t dt + \frac{1}{2}V_{SS}(dS)^2\right)e^{-rt} \\ &\qquad -rVe^{-rt}dt \\ &= \left(\frac{1}{2}V_{SS}\sigma^2S^2 + V_S rS + V_t - rV\right)e^{-rt}dt \\ &\qquad +V_S\sigma Se^{-rt}dw, \end{aligned}$$

where $(dw \cdot dw) = dt$, $(dt \cdot dw) = 0$ and $(dt \cdot dt) = 0$ (Merton [1992, p122–123]).

- However, $V^\dagger = Ve^{-rt}$ is a martingale in the risk-neutral world by construction, so there is no drift term. Thus, we deduce that

$$\frac{1}{2}V_{SS}\sigma^2S^2 + V_S rS + V_t - rV = 0.$$

Given the boundary conditions, we may solve this (Black-Scholes) PDE to find the option value $V(S(t),t)$. Different processes for dS yield different PDEs. We now value the product call.

• *END OF RISK−NEUTRAL PRICING REVIEW* •

Story: 1. Said he wasn't interested because the position paid too much. 2. While I was on a long-distance phone call, the applicant took out a copy of Penthouse, and looked through the photos only, stopping longest at the centerfold.

Interview Horror Stories from Recruiters
Reprinted by kind permission of *MBA Style Magazine*
©1996–2009 MBA Style Magazine, `www.mbastyle.com`

The time-t value of the European-style product call expiring at time-T is simply its discounted expected payoff in a risk-neutral world:

$$V(S_1(t), S_2(t), t) = e^{-r_{US}(T-t)}\, E^*\{\max[S_1(t)S_2(t) - X|\Omega_t]\},$$

where E^* denotes expectation taken with respect to the risk-neutral probability measure from the US perspective, and Ω_t is the time-t information set. We could work this out directly (it would be a double integral with respect to the two Brownian motions), but let us instead use the PDE approach.

Given the nature of the product call, I am going to guess that the solution is a function of only two variables, not three: $V(S_1, S_2, t) = \kappa H(\eta, t)$ for some constant κ and $\eta = S_1 \cdot S_2$ (see analogous guess in Wilmott [1998, p155]). I will need to use Itô's Lemma soon so I will now work out all the partial derivatives for the change of variables:

$$\begin{aligned}
\frac{\partial}{\partial S_1} &= \frac{\partial \eta}{\partial S_1}\frac{\partial}{\partial \eta} = S_2\frac{\partial}{\partial \eta}\\
\frac{\partial}{\partial S_2} &= \frac{\partial \eta}{\partial S_2}\frac{\partial}{\partial \eta} = S_1\frac{\partial}{\partial \eta}\\
\frac{\partial^2}{\partial S_1^2} &= S_2\frac{\partial \eta}{\partial S_1}\frac{\partial^2}{\partial \eta^2} = S_2^2\frac{\partial^2}{\partial \eta^2}\\
\frac{\partial^2}{\partial S_2^2} &= S_1\frac{\partial \eta}{\partial S_2}\frac{\partial^2}{\partial \eta^2} = S_1^2\frac{\partial^2}{\partial \eta^2}\\
\frac{\partial^2}{\partial S_1 \partial S_2} &= \frac{\partial}{\partial \eta} + S_2\frac{\partial \eta}{\partial S_2}\frac{\partial^2}{\partial \eta^2} = \frac{\partial}{\partial \eta} + S_1 S_2\frac{\partial^2}{\partial \eta^2},
\end{aligned}$$

and $\frac{\partial}{\partial t}$ is unchanged.

From our risk-neutral pricing review, we know $Ve^{-r_{US}t}$ is a martingale in the risk-neutral world, so it has no time trend. We need only find the coefficient of dt in $d[Ve^{-r_{US}t}]$ and equate it to zero. There are two Brownian motions, so we need the two dimensional Itô's Lemma (Merton [1992, p122]; Hull [1997,

p304]), and $d[Ve^{-r_{US}t}]$ is itself a geometric Brownian motion (GBM):

$$\begin{aligned}
& d[Ve^{-r_{US}t}]\\
&= -r_{US}Ve^{-r_{US}t}dt + e^{-r_{US}t}dV\\
&\overset{\text{Itô}}{=} -r_{US}Ve^{-r_{US}t}dt + e^{-r_{US}t} \times \left(\frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S_1}dS_1 + \frac{\partial V}{\partial S_2}dS_2\right.\\
&+ \left.\frac{1}{2}\frac{\partial^2 V}{\partial S_1^2}(dS_1)^2 + \frac{1}{2}\frac{\partial^2 V}{\partial S_2^2}(dS_2)^2 + \frac{\partial^2 V}{\partial S_1 \partial S_2}(dS_1 \cdot dS_2)\right)\\
&= e^{-r_{US}t} \times \left\{\left[-r_{US}V + \frac{\partial V}{\partial t} + \frac{\partial V}{\partial S_1}r_1 S_1 + \frac{\partial V}{\partial S_2}r_2 S_2\right.\right.\\
&+ \left.\frac{1}{2}\frac{\partial^2 V}{\partial S_1^2}\sigma_1^2 S_1^2 + \frac{1}{2}\frac{\partial^2 V}{\partial S_2^2}\sigma_2^2 S_2^2 + \frac{\partial^2 V}{\partial S_1 \partial S_2}\rho\sigma_1\sigma_2 S_1 S_2\right] dt\\
&+ \left.\left[\frac{\partial V}{\partial S_1}\sigma_1 S_1 dw_1 + \frac{\partial V}{\partial S_2}\sigma_2 S_2 dw_2\right]\right\}, \text{ which is a GBM,}
\end{aligned}$$

where we used the earlier definitions of dS_1, dS_2, and so on. We now take the time trend coefficient of dt, equate it to zero, use the change of variables $V(S_1, S_2, t) = \kappa H(\eta, t)$, where $\eta = S_1 S_2$, and drop the common terms $e^{-r_{US}t}\kappa$:

$$\begin{aligned}
-r_{US}H + H_t + r_1 S_1 S_2 H_\eta + r_2 S_1 S_2 H_\eta + \frac{1}{2}\sigma_1^2 S_1^2 S_2^2 H_{\eta\eta}\\
+\frac{1}{2}\sigma_2^2 S_1^2 S_2^2 H_{\eta\eta} + S_1 S_2 \sigma_1 \sigma_2 \rho [H_\eta + S_1 S_2 H_{\eta\eta}] = 0
\end{aligned}$$

Now collect terms and use $\eta = S_1 S_2$:

$$H_t + \eta H_\eta (r_1 + r_2 + \rho\sigma_1\sigma_2) + \frac{1}{2}(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)\eta^2 H_{\eta\eta} - r_{US}H = 0$$

Now plug in $r_1 = r_{US} - r_{JP}$ and $r_2 = r_{JP} - q - \rho\sigma_1\sigma_2$, and let $\sigma' \equiv \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}$ to deduce

$$H_t + \eta H_\eta (r_{US} - q) + \frac{1}{2}\sigma'^2 \eta^2 H_{\eta\eta} - r_{US}H = 0.$$

This PDE is the regular Black-Scholes PDE with continuous dividends and special volatility σ'. Recalling our definition of η, we get a "similarity solution" by using what we already know about the Black-Scholes solution to this PDE:

$$\begin{aligned}
c(t) &= S_1(t)S_2(t)e^{-q(T-t)}N(d_1) - e^{-r_{US}(T-t)}XN(d_2), \text{ where}\\
d_1 &= \frac{\ln\left(\frac{S_1(t)S_2(t)}{X}\right) + (r_{US} - q + \frac{1}{2}\sigma'^2)(T-t)}{\sigma'\sqrt{T-t}},\\
d_2 &= d_1 - \sigma'\sqrt{T-t}, \text{ and}\\
\sigma' &= \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}
\end{aligned}$$

Reassuringly, this is identical to equation (5.35) in Haug (2007, p227).

> **Advice:** *As an interviewer, I find telephone interviews difficult. Body language and nuances of voice are lost. Be sure to use a landline not a mobile/cell, hold the speaker close to your mouth, and speak a bit louder than usual.*

Answer 2.55: An Asian option is an average rate option. The underlying is a time series average of prices. Changes in average prices are much less volatile than changes in consecutive prices. Other things being equal, this lower volatility makes Asian options less expensive than plain vanilla options.

Answer 2.56: If the riskless rate is positive, and there are no dividends, early exercise is not optimal for an American-style call, and the European and American call have the same value. If the riskless rate is zero, then there is no incentive for early exercise of an American-style put. In this case, the European and American put have the same value.

Of course, that's little consolation to you if you are short an American-style option, a retail investor decides to exercise non-optimally, and you are assigned.

Answer 2.57: The one thing to watch out for here is that there are $(N+1)$ terms in the summation in both cases.

In the case of a recombining tree the answer is $1+2+3+4+\cdots+N+(N+1) = \sum_{i=1}^{N+1} i = \frac{(N+1)(N+2)}{2}$ (using the answer to Question 1.2 but with $(N+1)$ in place of n).

In the case of the non-recombining tree, there are $2^0+2^1+2^2+\cdots+2^N = \sum_{i=0}^{N} 2^i = 2^{N+1}-1$ nodes. There is a simple trick to get this last result if you cannot recall it. Let $S = \sum_{i=0}^{N} 2^i$, then multiply both sides by 2: $2S = \sum_{i=1}^{N+1} 2^i$. The RHS is just $S-1+2^{N+1}$, so you have $2S = S-1+2^{N+1}$, and you can solve directly for S.

Just out of interest, let me mention that this question came from a big name investment bank, and the candidate who answered it got both answers wrong. He did not realize it at the time and was not told he was wrong by the interviewer. A simple manual check of the formula compared to a diagram in the case $N=1$ or $N=2$ would have been enough to show him he was wrong!

Answer 2.58: The derivatives will give more leverage than the stock, so you should use a derivative. If you can only go long, then buying a call option is expensive because you are paying for the embedded downside protection of a put option via put-call parity type arguments (see Crack [2009, section 3.6]). With non-stochastic interest rates, the forward and futures prices are the same (Cox and Rubinstein [1985, p62]). It comes down to a question of margin/collateral. If the forward contract requires no collateral, and the futures contract requires a margin deposit, then the forward contract will provide more "bang for your buck."

Answer 2.59: There is not enough information to pinpoint a value; The interviewer wants a directional answer only and an explanation. So, if you are peeking here for advice, then go away and figure out only whether the call price should rise or fall.

Most stocks have positive betas, and the call option (as a leveraged investment in the stock) will therefore have a very large positive beta and a high positive expected return. For example, suppose a \$50 stock has a beta of $\beta_S = 1.10$. Then an at-the-money six-month European call option on the stock (assume $r = 0.05$, $\sigma = 0.30$) has a beta of $\beta_c = 6.719837$ under Black-Scholes assumptions.[63] This is an "instantaneous beta" because as soon as the stock price changes, the degree of leverage in the call changes and the call's beta changes (but it will still be higher than the stock's beta). So, the expected return on the option is positive and tomorrow's expected price is higher than today's.

Answer 2.60: Now, how do we reconcile Answer 2.59 (call price expected to rise) with negative theta (i.e., time decay)? An option's theta is the sensitivity of the option price to the passage of time holding all else constant. You cannot look at this in isolation because all else is not held constant over the next 24 hours. To reconcile the positive expected return on the option with the negative theta, we need a formula that uses a total differential.

If we were working with deterministic functions, we would simply write

$$dc = \frac{\partial c}{\partial t}dt + \frac{\partial c}{\partial S}dS = \Theta dt + \Delta dS. \quad (\dagger)$$

We could then talk about how delta is positive (i.e., $\Delta \equiv \frac{\partial c}{\partial S} > 0$) and how the expected value of dS is positive (because stock price is expected to rise on average) and how this outweighs the negative time decay term (i.e., the negative theta: $\Theta \equiv \frac{\partial c}{\partial t} < 0$) and so on. This is, however, quite wrong here!

Continuing with the numerical at-the-money call example from Answer 2.59, on the stock with beta 1.10, and using a 10% expected return for a CAPM Market portfolio (see notes at the end of this section), I have the following numbers: $\Theta = -5.357262$, $dt = \frac{1}{365}$, $\Delta = 0.588589$, $E(dS) \approx 0.013679$ (see notes at end of this question). So, $\Delta E(dS) \approx \$0.008051$, but $\Theta dt = -\$0.014677$, so the total differential above (i.e., equation $(\dagger)$) would give $E(dc) = -\$0.006626$ which contradicts that fact that the expected return on the call is positive.

What is missing is that the total differential above (i.e., equation $(\dagger)$) applies only to deterministic functions. The call price is stochastic, driven by the stochastic S. We cannot use the stated total differential because it has a term

[63] The relationship is $\beta_c = \Omega\beta_S$, where $\Omega \equiv \frac{N(d_1)\cdot S}{c} = \frac{\Delta\cdot S}{c}$ is the elasticity of the call price with respect to the stock price (Cox and Rubinstein [1985, p190]). In this example we have $c = 4.817438$, $S = 50$, and $N(d_1) = \Delta = 0.588589$.

missing! We need Itô's Lemma:

$$
\begin{aligned}
dc &= \frac{\partial c}{\partial t}dt + \frac{\partial c}{\partial S}dS + \frac{1}{2}\frac{\partial^2 c}{\partial S^2}(dS)^2 \\
&= \Theta dt + \Delta dS + \frac{1}{2}\Gamma(dS)^2,
\end{aligned}
$$

where $\Gamma \equiv \frac{\partial^2 c}{\partial S^2}$. Continuing the numerical example above, we have $\Gamma = 0.036681$, and $E[(dS)^2] \approx \$0.617039$ (see notes at end of this question), so then we get

$$
\begin{aligned}
E(dc) &= \Theta dt + \Delta E(dS) + \frac{1}{2}\Gamma E\left[(dS)^2\right] \\
&\approx \left(-5.357262 \cdot \frac{1}{365}\right) + (0.588589 \cdot 0.013679) \\
&\qquad + \left(\frac{1}{2} \cdot 0.036681 \cdot 0.617039\right) \\
&= -0.014677 + 0.008051 + 0.011317 \\
&= \$0.004691.
\end{aligned}
$$

This is roughly a 10 bps increase in value from $c = \$4.817438$. That is, the time decay contributes a negative component to the expected change in the call price, but the expected values of the stochastic terms contribute positive values that more than offset the time decay.

In simple terms, the contributions to the expected change in call price are: a negative term for the time decay, a positive term for the call's delta and the positive expected return on the stock, and a positive return for the call's gamma and the volatility of the stock.

Notes for readers who want more details: My numerical example is given to make the analysis concrete and to show relative magnitudes; I have given many decimal places so you can reproduce it. No interviewer expects this sort of detail. The numerical example uses $dt = \frac{1}{365}$ but this infinitesimal notation is not strictly correct for a non-infinitesimal time step. The same can be said of using the notation dS and dc for non-infinitesimal moves in the stock and call prices, respectively. I am simply going to ignore this loose use of notation and continue. If you are not happy with that then you are not ready to interview with a finance firm. I gave numbers for $E(dS)$ and $E[(dS)^2]$ without saying where they came from. Here are the minimum details: If μ is expected return on the stock such that $E[S(t+dt)] = S(t)e^{\mu dt}$, then in a Black-Scholes world we may write (loosely) $S(t+dt) = S(t)e^{(\mu - \frac{1}{2}\sigma^2)dt + \sigma\sqrt{dt}\cdot\epsilon}$, where $\epsilon \sim \mathcal{N}(0,1)$. In this case, $dS = S(t+dt) - S(t)$, and we can use the properties of Normal and Lognormal distributions (see Question 4.22) to deduce that $E(dS) = S(t)(e^{\mu dt} - 1)$ and $E[(dS)^2] = S^2(t)(e^{2(\mu + \frac{1}{2}\sigma^2)dt} - 2e^{\mu dt} + 1)$. You can also reproduce the value of $E(dc)$ almost exactly by using the CAPM information in the answer. Note however that the $E(dc)$ number implied by

the CAPM will match the $E(dc)$ value calculated using Itô's Lemma only in the limit as $dt \to 0$.[64] If you want to use the CAPM then you should use what approximates an instantaneous CAPM (a CAPM defined only over the time step dt). I used $dt = \frac{1}{365}$, $r = 0.05$, $Rf = e^{r\cdot dt} - 1$, $r_M = \ln(1 + 0.10)$, $E(R_M) = e^{r_M \cdot dt} - 1$, $\mu_{S,simple} = Rf + \beta_S(E(R_M) - Rf)$, $\mu = \frac{1}{dt}\ln(1 + \mu_{S,simple})$, $\beta_c = \Delta \cdot S \cdot \beta_S / c$, $\mu_{c,simple} = Rf + \beta_c(E(R_M) - Rf)$, $E(dc)_{CAPM} = c \cdot \mu_{c,simple}$.

[64] This is because the CAPM relationship between β_c and β_S given in Footnote 63 is only true in an instantaneous CAPM, and this holds only for an infinitesimal time step. If I look at the ratio of the $E(dc)$ from the CAPM argument to that from the Itô's Lemma argument, I get 97.6508% for $dt = \frac{10}{365}$, 99.7603% for $dt = \frac{1}{365}$, 99.9760% for $dt = \frac{0.1}{365}$, 99.9976% for $dt = \frac{0.01}{365}$, and 99.9998% for $dt = \frac{0.001}{365}$.