The Mathematics of Your Next Family Reunion

by

Jeffrey S. Rosenthal

(October, 2011.)

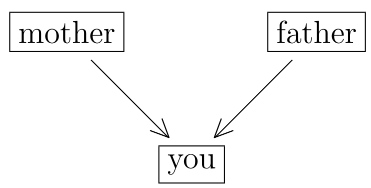
– *I am your father's brother's nephew's cousin's former roommate.*
– *What does that make us?*
– *Absolutely nothing!*

(Lord Dark Helmet and Lone Starr, discussing family relationships in the Mel Brooks movie *Spaceballs*.)

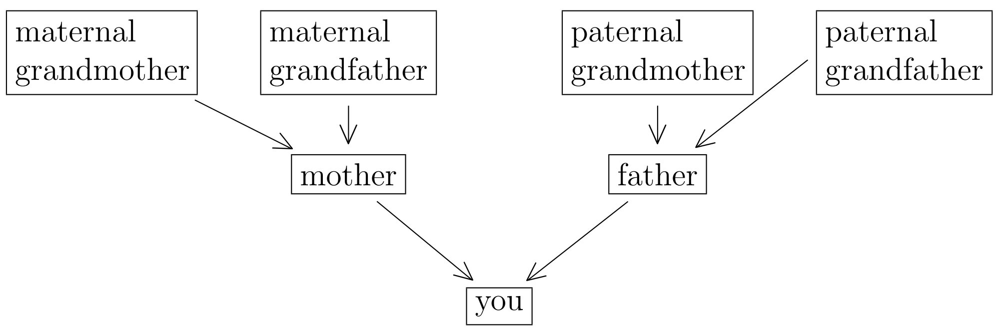
Keeping track of family relations can be difficult. If Edna marries your mother's uncle Charlie, what should you call her? If your father's cousin's daughter just had a baby boy, how should you two be introduced? Who is your "great great aunt", and how can you find your "first cousin twice removed"? Fortunately, a bit of mathematical logic can clarify who should be called what, and why – and even measure the degree of genetic similarity between different relatives.

Ancestor Lineage

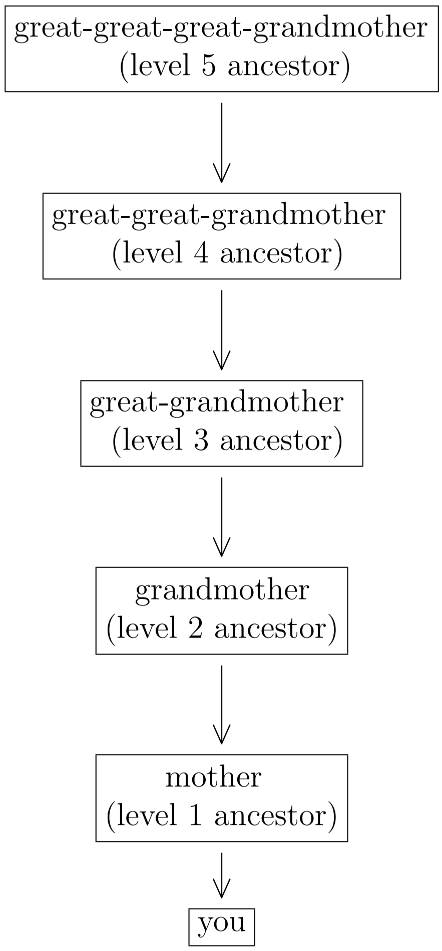
To begin at the beginning (well, *your* beginning, anyway), you surely had two parents, a mother and father:



Continuing backwards, they each had two parents, giving you a total of four grandparents:



Going back still further, each of your ancestors in turn had two parents, indicated by prepending an extra "great" each time. For example, your maternal lineage is:

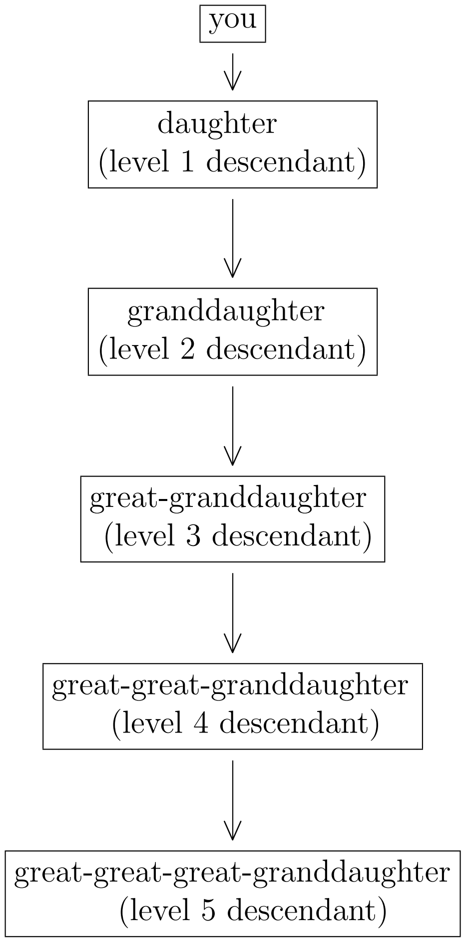


and so on (and similarly for "fathers" instead of "mothers" at any level).

Since each ancestor has two parents (one mother and one father), you have a total of 2^n ancestors at level n: two parents, four grandparents, eight great-grandparents, sixteen great-great-grandparents, etc. Summing up, you have a total of $2 + 2^2 + 2^3 + \ldots + 2^n = 2^{n+1} - 2$ ancestors of level n or lower; for example, your total number of parents and grandparents and great-grandparents combined is $2^{3+1} - 2 = 16 - 2 = 14$. In short, your ancestors form a perfect binary tree – simplicity itself.

Descendant Legacy

If you have children yourself, then *their* children are your grandchildren, and your grandchildren's children are your great-grandchildren, and so on:



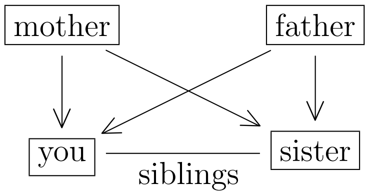
(and similarly for "son" instead of "daughter" at any level).

Unlike with ancestors, there is no simple formula for your number of descendants. Rather, you have to count up all of your children, and all of their children, and so on. For example, even if you have five children, it is possible that none of them will have children of their own, in which case your number of grandchildren will be zero. On the other hand, if they each have five children of their own, then you will have twenty-five grandchildren – a lot more.

Sideways, March!

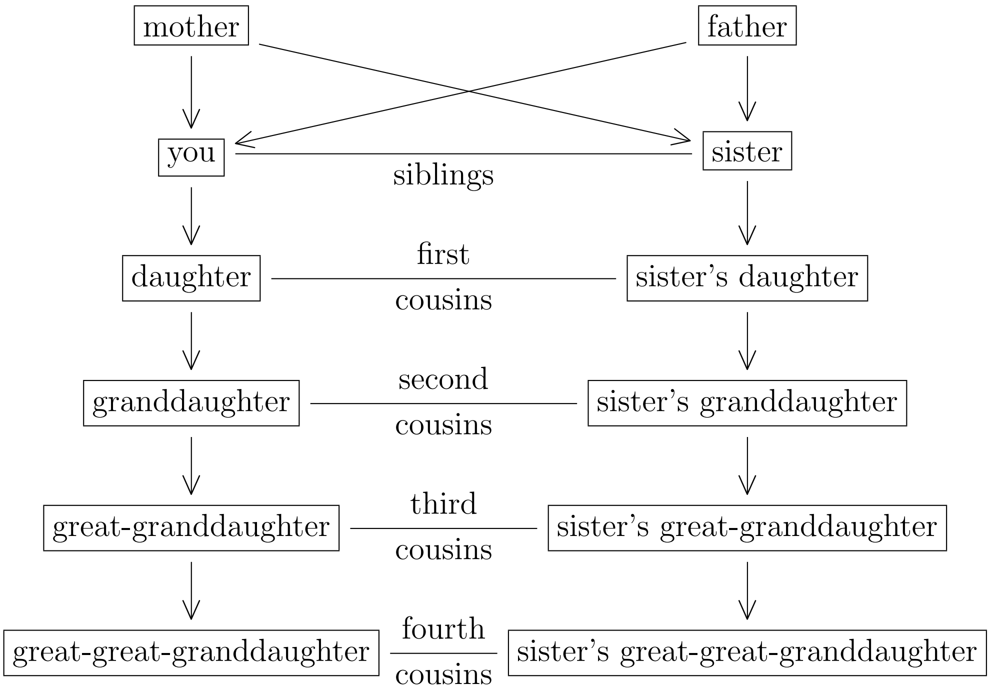
When people have more than one child, this fattens the family tree, creating new relationships like *sister* and *niece* and *great-aunt* and more.

For starters, if your parents have additional children besides you, then they are of course your siblings, i.e. your sisters and brothers:



(Here, and throughout, relationships to "you" are written within the boxes, and relationships between other pairs of individuals are indicated by connecting lines.)

If you and your siblings each have children, then those children are first-cousins of each other. Then, if the two first-cousins each have children, then *those* children are second-cousins of each other; and *their* children are third-cousins, and so on:

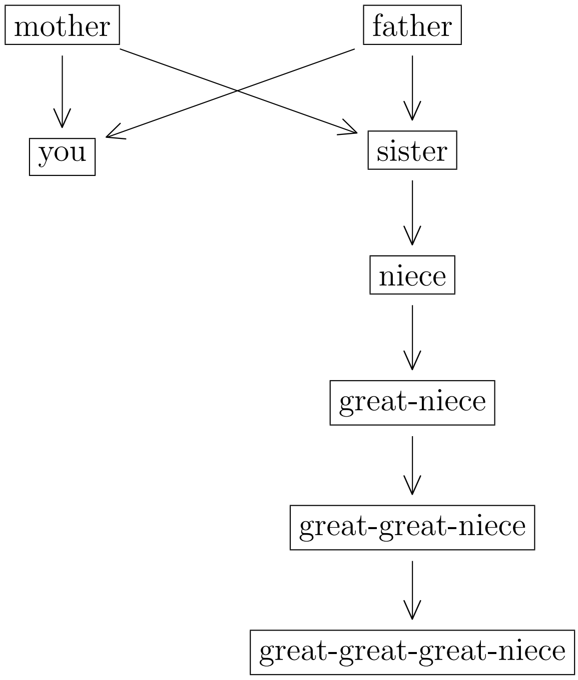


(and similarly for "son" instead of "daughter" at any level).

In general, n-level cousins share two $(n+1)$-level ancestors (but no n-level ancestors). Thus, first-cousins share two grandparents (but no parents), and second-cousins share two great-grandparents (but no grandparents), and so on.

It follows that if A and B are n-level cousins, then A's child and B's child are $(n+1)$-level cousins. Thus, children of first-cousins are second-cousins, and children of second-cousins are third-cousins, and so on. In fact, if we regard siblings as 0-level cousins, then this reasoning applies to siblings too: children of 0-level cousins (i.e., siblings) are themselves first-cousins.

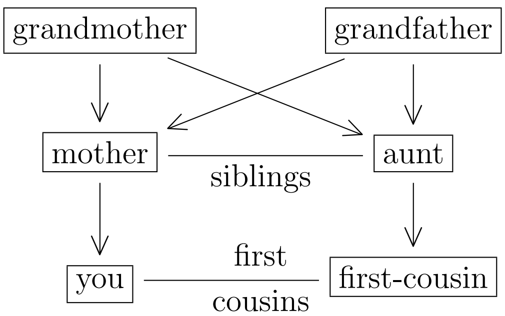
Finally, your sibling's child is your niece (or nephew, if male), and *their* child is your great-niece (or great-nephew), and so on:



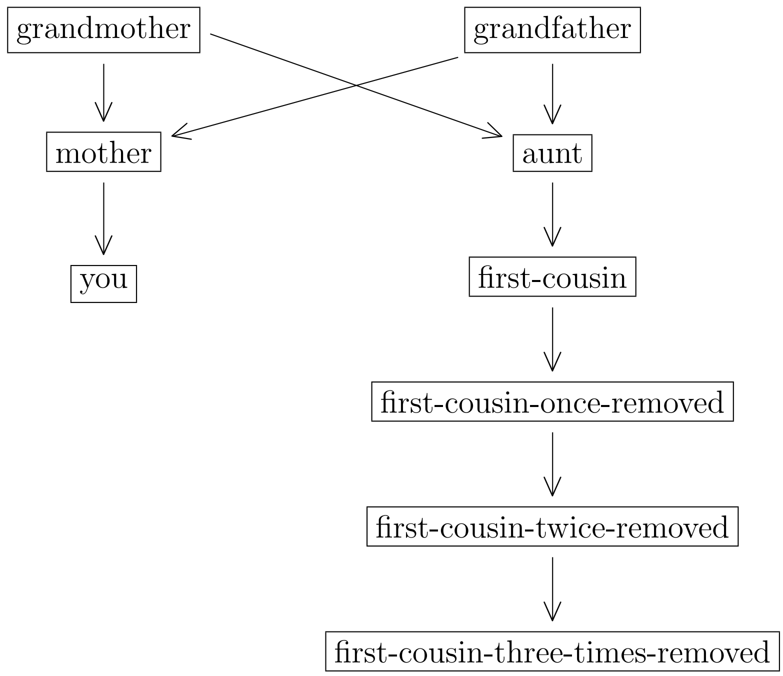
(and similarly for “nephew” instead of “niece” at any level).

Cry Uncle

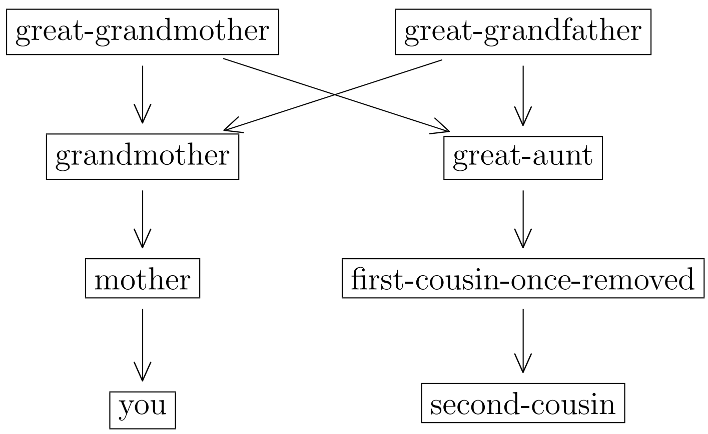
So now we know where your descendants’ cousins come from. To see where *your* cousins come from, we have to move up to your parents’ level. Your parents’ siblings are your aunts and uncles, and their children are your first-cousins (since you and they share the same grandparents, but not the same parents):



If your cousins have children, then what are they to you? Well, children of your first-cousin are called your “first-cousins-once-removed”, and *their* children are your “first-cousins-twice-removed”, and so on:

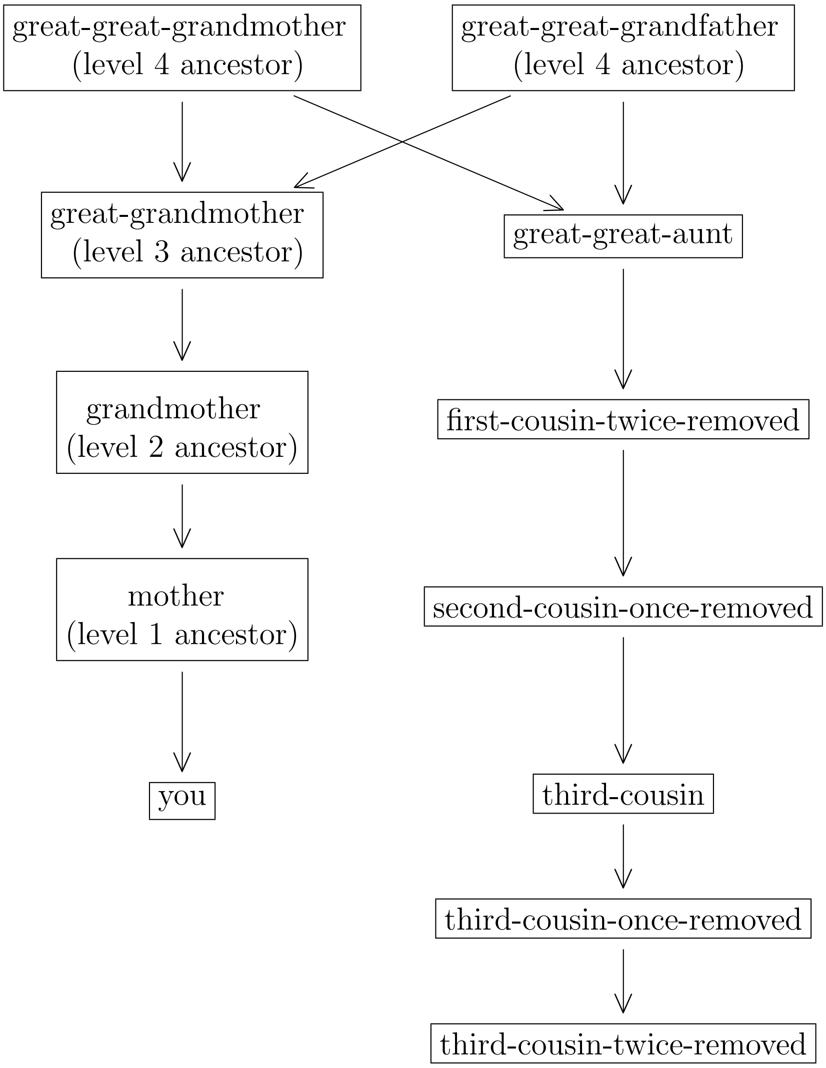


To see where your second-cousins come from, we have to move one more level up. Your grandparents' siblings are your great-aunts and great-uncles. So their children, i.e. your parents' cousins, are your first-cousins-once-removed. And *their* children are your second-cousins:



The same pattern continues upwards for all earlier generations. Once again, your n^{th} cousins share your $(n+1)$-level ancestors, but not your n^{th}-level ancestors. Siblings of your n^{th}-level ancestors are your great-…-great aunts and great-…-great uncles, where "great" is repeated $n-1$ times. Furthermore, the n^{th} cousins of your m^{th}-level ancestors, and also the m^{th}-level descendants of your n^{th} cousins, are your n^{th} cousins m times removed.

For example, with $n = 3$ and $m = 2$, this says that your grandparents' third-cousins are your third-cousins-twice-removed, and your third-cousins' grandchildren are also your third-cousins-twice-removed. Tracing back to $n = 3$ gives:



In this diagram, your third-cousin ($n = 3$) shares two of your great-great-grandparents (level $n + 1 = 4$ ancestors) but none of your great-grandparents (level $n = 3$ ancestors). Your great-great-aunt is a sibling of your great-grandmother ($n = 3$). Your second-cousin-once-removed achieved that designation by being the second cousin ($n = 2$) of your mother (level $m = 1$ ancestor), while your third-cousin-once-removed achieved that designation by being the daughter (level $m = 1$ descendant) of your third cousin ($n = 3$). Tricky!

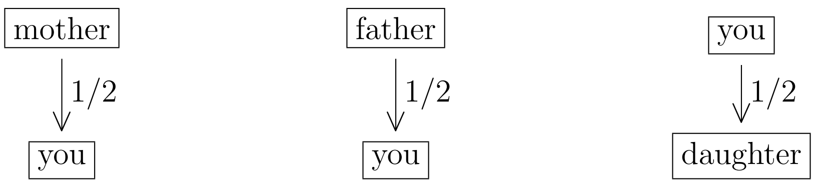
Thicker Than Water

One of the reasons we care about family trees is because of a sense that certain family relations are "more related" to us, and should be assisted and protected and loved on that basis. This attitude presumably has an evolutionary basis: our genes survived through the ages because our ancestors made efforts to help them survive by caring not only for themselves, but for their close relatives too. Indeed, there is an ancient Bedouin Arab saying, "I against my brother, my brothers and me against my cousins, then my cousins and I against strangers", which nicely illustrates the philosophy of caring most for those who are genetically closest to us.

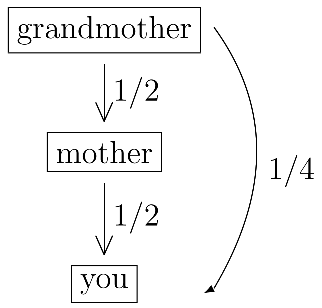
This raises the question of just how similar our relatives' genes are to our own. Well, first of all, about 99.9% of our genetic material is common to all humans (yes, even your in-laws), and indeed is what makes us human. Furthermore, some people may share other genes with us just by chance; for example, if I meet a stranger whose eyes are brown just like mine are, that does not necessarily establish that we are close relatives. In addition, there is lots of *randomness* in how genes are passed on (each individual gets half of their genetic material from their mother and half from their father, but which bits come from which parent is chosen at random and cannot be predicted), so we cannot draw precise conclusions with certainty.

To deal with all of this, we assign to each pair of individuals a *relatedness coefficient* which represents the *expected fraction* (i.e., the fraction on average) of their genes which are *forced* to be identical by virtue of their family relationship. This approach averages out all of the randomness, while focusing on genetic similarities specifically due to family connections.

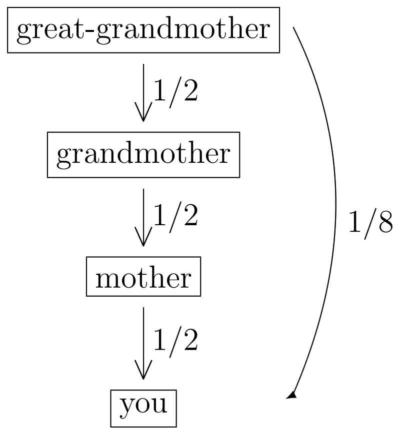
According to this definition, strangers have a relatedness of zero (the smallest possible value). By contrast, your relatedness with *yourself* is one (the largest possible value). Other relatedness coefficients fall between these two extremes. For example, your relatedness with your mother is 1/2, since you obtain half of your genetic material from her. And your relatedness with your father is also 1/2. By the same reasoning, your relatedness with your child is again 1/2. So far so good:



Next consider your maternal grandmother. She gave half of her genes to your mother, and then your mother gave half of her genes to you. It is *possible* that the half you took is exactly the same as the half your grandmother gave. It is also *possible* that the half you took has no overlap at all with the half your grandmother gave. But on average, i.e. in *expectation*, exactly half of the genetic material you took from your mother originated from your maternal grandmother. So, your relatedness coefficient with your grandmother is one-half of one-half, i.e. $(1/2) \times (1/2)$, or 1/4:



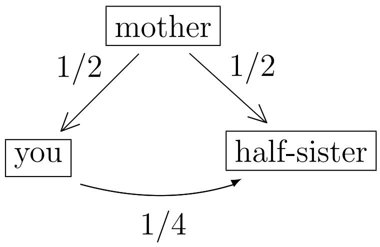
Continuing up the tree, your relatedness with your great-grandmother is one-half of one-half of one-half, i.e. $(1/2) \times (1/2) \times (1/2)$, or $1/8$:



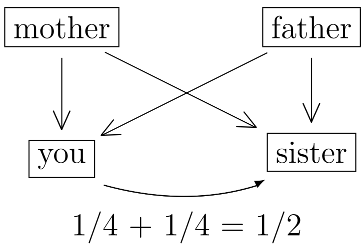
(and similarly for "father" instead of "mother" at any level). In general, your relatedness coefficient with your level-n ancestor is $1/2^n$.

By the same reasoning, your relatedness coefficient with your level-n descendant is also $1/2^n$. So, for example, your relatedness coefficient with your daughter is $1/2$; with your granddaughter is $1/4$; and with your great-granddaughter is $1/8$ (and similarly for "son" instead of "daughter").

For siblings, the situation is a little bit more complex. Consider first the case of two *half-siblings* (half-sisters or half-brothers), i.e. people who share just one parent. Since they each got half of their genetic material from that one shared parent, their relatedness coefficient is one-half of one-half, i.e. $1/4$:

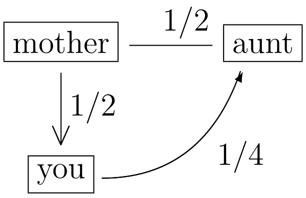


Regular (full) siblings similarly share 1/4 of their genetic material through their mother, but *also* share 1/4 of their genetic material through their father. This gives a total relatedness coefficient of $1/4 + 1/4 = 1/2$:



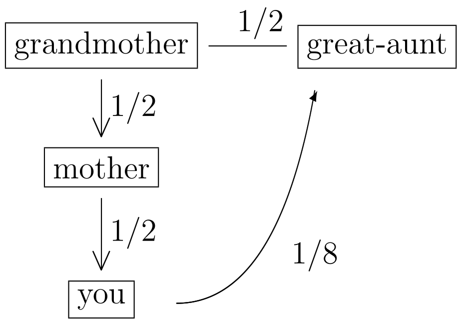
(One special case is *identical twins*, who have identical genes and thus a relatedness coefficient of one. But fraternal twins have relatedness coefficient 1/2, just like other siblings.)

Continuing onward, since your mother and aunt are siblings, they have relatedness coefficient 1/2. Meanwhile, you and your mother have relatedness coefficient 1/2. Putting this together, you and your aunt (or uncle) have relatedness coefficient $(1/2) \times (1/2) = 1/4$:



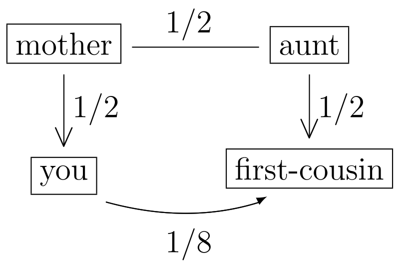
(and similarly with "aunt" replaced by "uncle"). And, your relatedness coefficient with your niece or nephew is also 1/4.

Taking it to the next level, your grandmother and your great-aunt are also siblings and hence also have relatedness coefficient 1/2. Since you have relatedness coefficient 1/4 with your grandmother, it follows that you have relatedness coefficient 1/8 with your great-aunt:

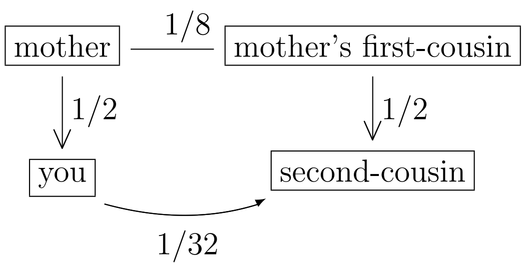


(and similarly with "great-aunt" replaced by "great-uncle"). Similarly, your relatedness coefficient with your great-niece or great-nephew is also 1/8.

Then, since your first-cousin has relatedness coefficient 1/2 with your aunt, who in turn has relatedness coefficient 1/4 with you, it follows that you and your first-cousin share relatedness coefficient 1/8:



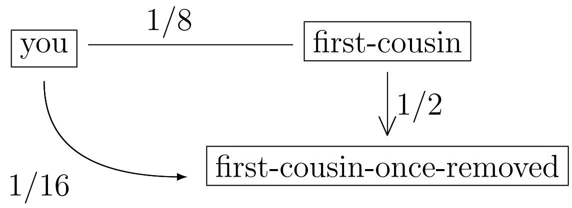
Now, since your mother and her first-cousin have relatedness coefficient 1/8, and since you have relatedness coefficient 1/2 with your mother, and since your mother's first-cousin has relatedness coefficient 1/2 with her own child (who is your second-cousin), it follows that your relatedness coefficient with your second-cousin is $(1/2) \times (1/8) \times (1/2) = 1/32$:



In general, switching to level-n cousins from level-$(n-1)$ cousins introduces two new factors of 1/2. Since $(1/2) \times (1/2) = 1/4$, this means that your relatedness coefficient with your level-n cousin is always 1/4 times your relatedness coefficient with your level-$(n-1)$ cousin. It follows that your relatedness coefficient with your level-n cousin is equal to $1/2^{2n+1}$.

So, your relatedness coefficient with your first cousin is 1/8; with your second cousin is 1/32; with your third cousin is 1/128; and so on.

What about first-cousins-once-removed, and all of that? Well, since you and your first-cousin have relatedness 1/8, and since your first-cousin and their child (your first-cousin-once-removed) have relatedness 1/2, it follows that you and your first-cousin-once-removed have relatedness coefficient $(1/8) \times (1/2) = 1/16$:



The pattern continues, with each new "removed" introducing an extra factor of 1/2 into the product. It follows that your relatedness coefficient with your n^{th} cousin, m times removed, is equal to $1/2^{2n+m+1}$. For example, your relatedness coefficient with your third cousin ($n = 3$) twice removed ($m = 2$) is equal to $1/2^{6+2+1} = 1/2^9 = 1/512$ – not very close at all.

We can summarise the relatedness coefficients of various relationships in a table:

Relationship to you	**relatedness coefficient**
yourself	1
identical twin	1
parent, child	1/2
grandparent, grandchild	1/4
great-grandparent, great-grandchild	1/8
n^{th} level ancestor or descendant	$1/2^n$
sibling (sister or brother)	1/2
half-sibling	1/4
aunt, uncle	1/4
niece, nephew	1/4
great-aunt, great-uncle	1/8
great-niece, great-nephew	1/8
first-cousin	1/8
first-cousin-once-removed	1/16
second-cousin	1/32
second-cousin-once-removed	1/64
third-cousin	1/128
n^{th} cousin	$1/2^{2n+1}$
n^{th} cousin, m times removed	$1/2^{2n+m+1}$
stranger	0

This table can be thought of as indicating your level of evolutionary imperative to protect and assist your various relatives. That perspective was nicely summarised by the early evolutionary biologist J.B.S. Haldane, when he was asked if he would give his life to save a drowning brother, and replied "No, but I would to save two brothers or eight cousins." He was merely observing that $2 \times (1/2) = 8 \times (1/8) = 1$, i.e. that two brothers, or eight cousins, are each "equal" (in evolutionary terms) to one copy of yourself.

So what about that saying, "I against my brother, my brothers and me against my cousins, then my cousins and I against strangers"? Well, in the context of relatedness coefficients, it corresponds to the observation that your relatedness coefficient is higher with yourself (1) than with your brother (1/2), higher with your brother (1/2) than with your first-cousin (1/8), and higher with your first-cousin (1/8) than with a stranger (0). That is:

$$1 \;>\; 1/2 \;>\; 1/8 \;>\; 0\,.$$

It seems that those Bedouins knew their inequalities well!

Families of All Shapes and Sizes

Of course, the evolutionary imperative associated with relatedness coefficients does not tell the whole story. You would (hopefully) protect your spouse over your second-cousin even though, strictly speaking, your relatedness coefficient with your spouse is zero (since you have no actual blood relationship). And, parents of adopted children should surely treat them just like biological children, despite the lack of true genetic connection.

Other family relationships can arise too. For example, if you marry, then your spouse's relations become your corresponding *in-law* relations – exactly as they are for your spouse, except with the suffix "in-law" appended. For example, your husband's father is your father-in-law, your husband's cousin is your cousin-in-law, and so on. The suffix "in-law" is also used for those who marry your relations – for example, your brother's wife is your sister-in-law. (One exception is that your aunt's husband gets to be called your uncle, even though he is "really" your uncle-in-law; and similarly your uncle's wife gets to be called your aunt.) These in-law rules can be combined – for example, a friend of mine cheerfully calls his wife's sister's husband his "brother-in-law-in-law". Of course, your genetic relatedness coefficient with your in-laws is zero, since your relationship is through marriage rather than actual blood lines.

If your father (say) has children with a partner other than your mother (usually, after a divorce and remarriage), then those children are your "half-siblings" (as already mentioned), with relatedness coefficient 1/4, i.e. half that of regular (full) siblings. This "halfness" then continues on, e.g. your half-sibling's children are your half-nieces and half-nephews, with relatedness coefficients $(1/2) \times (1/4) = 1/8$. Meanwhile, a woman who marries your father after your mother becomes your *step-mother* (or step-father, if the genders are reversed). Her relations become your corresponding step-relations – exactly as if your step-mother were your mother, except with the prefix "step" appended. For example, your father's second wife's brother is your step-uncle, and his children are your step-cousins, and so on. Your genetic relatedness coefficient with your step-relations is also zero, since again your relationship is through marriage not blood.

Further variations arise by considering other species with other mechanisms for genetic selection. For example, if a species propagates by *cloning*, i.e. making exact genetic copies, then all relations have identical genetic material, and hence a relatedness coefficient of one.

More subtly, many species of bees and ants are *haplodiploid*. This means that their males have half as much genetic material as their females. In fact, males have no father at all, but rather get all their genes from their mother. By contrast, females get half of their genes from their mother (by taking half of her genes, chosen at random, just like humans) and half of their genes from their father (by taking *all* of his genes, with no randomness at all). In such a system, the relatedness coefficient is no longer symmetric. Mothers and daughters still have a relatedness coefficient of 1/2, and mothers still have a relatedness coefficient of 1/2 with their sons. However, sons have a relatedness coefficient of 1 with their mothers (who contain *all* of their genetic material). Most interestingly, full sisters have a relatedness coefficient of 3/4, since they share *all* of their father's genes in addition to half of their mother's. So, since $3/4 > 1/2$, female bees and ants have a genetic imperative to assist their sisters even ahead of their own children!

Family relations can lead to unexpected surprises. At a recent large family reunion, I met a young man whom I did not know. After some discussion, we determined that my great-grandfather was the brother of his great-grandmother – making us third cousins. Furthermore, my great-grandmother was the sister of his great-grandfather, too. That is, three generations earlier, a brother-and-sister pair had married off with a sister-and-brother pair. This meant that he and I were third-cousins by each of two different paths – we were "double third-cousins"! It followed that our relatedness coefficient was *twice* that of usual third-cousins, i.e. equal to $2 \times (1/128) = 1/64$ – still not very close, but interesting nonetheless. I wish I had had the presence of mind to immediately say to him, "Pleased to meet you, double-third-cousin. I am honoured to share one sixty-fourth of your genes."

Jeffrey S. Rosenthal is a professor in the Department of Statistics at the University of Toronto, and the author of the popular book "Struck by Lightning: The Curious World of Probabilities" (Granta, 2008; www.probability.ca/sbl).
