The chain rule. Although the chain rule is most easily remembered in the form

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}, \qquad (*)$$

we need the equivalent formulation of $(*)$ in terms of functions to derive the substitution rule for integrals. Thus, write

$$y = g(x) \text{ and } z = F(y)$$

to obtain

$$\frac{d}{dx} F\big(g(x)\big) = F'\big(g(x)\big) \cdot g'(x).$$

Writing $F' = f$ and writing the statements about derivatives with respect to x or y as statements about integrals with respect to x or y, we get

$$F\big(g(x)\big) = \int f\big(g(x)\big) \cdot g'(x)\,dx$$
$$F\big(y\big) = \int f\big(y\big)\,dy.$$

This shows the significance of the dx and dy as a means of identifying the variable in the notation for integrals: when we write $y = g(x)$, the derived expression $dy = g'(x)\,dx$ gives a purely notational way to include **exactly** the factor required by the chain rule.

Using the substitution rule. In Section 7.3, we will meet an example where writing $y = g(x)$ turns an integral with respect to y that we don't know how to evaluate into an integral with respect to x that is known. Such applications are *rare*. Most of the time, the substitution formula is used to identify an expression in an integral with respect to x that can play the role of y. This allows the desired integral to be found in the form $F\big(g(x)\big)$.

Examples. The following will be done at the blackboard.

$$(\#1) \qquad \int x\big(x^2 - 1\big)^{99}\,dx$$

$$(\#11) \qquad \int x^3\sqrt{2 + x^4}\,dx$$

$$(\#27) \qquad \int \cos^4 x \sin x \, dx$$

$$(\#37) \qquad \int \frac{\left(\ln x\right)^2}{x} \, dx$$

A paraphrase of the substitution rule. To write an integral with respect to x as a function of some expression $y = g(x)$, you need to (1) extract a factor of $g'(x)$ from the integrand, and (2) express *everything else* as a function of y. Some common examples are: (1) if the integrand is a polynomial all of whose terms have odd degrees, take $y = x^2$ with $dy = 2x\,dx$; (2) $\sin x$ times any power of $\cos x$, take $y = \cos x$ with $dy = -\sin x\,dx$; (3) a function of $ax + b$, take $y = ax + b$ with $dy = a\,dx$.

Definite integrals. The usual evaluation of

$$\int_a^b h(x)\,dx,$$

calls for identifying $H(x)$ with $H' = h$, and then, the integral is $H(b) - H(a)$. In other words, a and b are

values that are substituted for x. If

$$h(x) = f\big(g(x)\big) \cdot g'(x),$$

then $H(x) = F\big(g(x)\big)$, where $F'(y) = f(y)$. A definite integral with this integrand will have a value of the form $F(d) - F(c)$, with the quantities c and d obtained as the values of the function g at the original limits of integration. These numbers are the values of y that correspond to the original values of x that were the original limits of integration.

A definite integral evaluates to a number. Any variables that appear in the expression serve only to allow the fundamental theorem to be applied. If a substitution $y = g(x)$ is used to evaluate the integral, then it is not necessary to evaluate the indefinite y integral, substitute $y = g(x)$ in that formula, and evaluate at particular values of x. Instead, you can immediately evaluate the function $g(x)$ at the limits of the x integral to get the limits of the y integral. The resulting definite integral in y will give the desired answer. All history of the original variable x and the substitution $y = g(x)$ can now be erased. They have done their job.

Areas between curves. If $f(x) \geq g(x) \geq 0$ for $a \leq x \leq b$, the region below $y = f(x)$ and above $y = g(x)$ and be found by *removing* the area under $y = g(x)$ from the whole region under $y = f(x)$ and above the x axis. The desired area *appears to be* just the difference of the integrals of $f(x)$ and $g(x)$, which is the integral of the difference $f(x) - g(x)$. There is a small problem with this argument: our understanding of area is based on formulas for areas of *polygons* that are extended by a limit process. Since limits need not exist, any extension of our definition of area must check that we have not exceeded the resources of our limit argument. Everything works in this case, and we wind up with a more general definition of area than was used originally.

Note that only the position of $f(x)$ and $g(x)$ relative to each other, not their location relative to the x axis, is relevant in the final formula.

Setting up the integral. If you are just asked to find the area between $y = 4x^2$ and $y = x^2 + 3$ (as in exercise 6.1.11), it is clear what f and g are, though not necessarily which is which, but the numbers a and b

are not evident. The usual way to find the missing information is with a picture — and I just happen to have one. All regions bounded by the given curves except the one in our picture are unbounded, so the bounded region must have been intended if a numerical answer is expected. For this region

$$4x^2 \leq y \leq x^2 + 3$$

and this implies $4x^2 \leq x^2 + 3$, which holds *exactly* for $-1 \leq x \leq 1$. Thus $a = -1$ and $b = 1$. In general, when no vertical sides of the region are given, the endpoints are found by solving $f(x) = g(x)$. A complete analysis requires determining which of the given functions is $f(x)$ — the top of the region — and which is $g(x)$ — the bottom. If you accidentally reverse these choices, the integral is replaced by its negative, so you can immediately replace your answer with the correct one — but you should check that this is the correct explanation, since errors in calculation could lead to totally irrelevant answers.

When the given curves cross several times, it is no longer clear what it means to be *between the curves*.

Rather than assume that the mathematics has an obligation to give a clear answer because such simple words were used to state the question, you should insist that all descriptions be clear enough to be directly translated into mathematical expressions of the form $g(x) \leq y \leq f(x)$ for $a \leq x \leq b$, as required by our formula for area. There are two conflicting interpretations: **signed area** and **total area**. In the first interpretation, you compute a simple integral and declare that it counts some areas positively and others negatively. This is a lazy way out favored by many mathematicians. Unfortunately, the textbook prefers to form the sum of the positive areas of all parts of the largest region consistent with the description. We call this the "total area".

Signed Areas. The integral of a **positive** function is identified with the area under the graph of that function and above the x axis. The definition of the integral makes sense even if the function is sometimes negative. Wherever $y = f(x)$ is above the axis, the integral gives the area below the curve, and wherever the graph is below the axis, the integral gives the neg-

ative of the area above the curve. For the area between curves, the result is easier. $\int f(x) - g(x)\,dx$ gives the area between the curves when $f(x) > g(x)$ and the negative of the area when $f(x) < g(x)$. The integral will always calculate an *algebraic sum* of areas, but some regions behave as if they have negative area. In multivariable calculus, this will be explained in terms of whether the description leads you to walk around the boundary of the region so that the inside is to your left (the positive case) or right (the negative case). For the time being, only simple examples will be used where it is easy identify the integral that gives the positive answer.

Total area. If you want to get the total area between curves, you need to compute

$$\int_a^b \left| f(x) - g(x) \right| \, dx.$$

If both $f(x)$ and $g(x)$ are continuous, the integrand is continuous, so the integral certainly exists. However, few differentiation formulas contain absolute values.

The usual way to calculate such an integral is to use

$$\int_a^c h(x)\,dx = \int_a^b h(x)\,dx + \int_b^c h(x)\,dx$$

to break the integral into a sum of expressions in which the integrand does not change sign. Then use

$$\left|h(x)\right| = \begin{cases} h(x) & h(x) \geq 0 \\ -h(x) & h(x) \leq 0 \end{cases}$$

to replace the summands by integrals suitable for the methods of elementary calculus.

I have a picture of exercise 6.1.23 to guide our solution of this exercise.