MATHEMATICS AS COMMON SENSE:
DERIVING THE BASIC TRIG FUNCTIONS

Clyde Greeno
The MALEI Mathematics Institute
P.O. Box 54845 Tulsa, OK 74155
greeno@mathematicsinstitute.org

Developmental Vs. Elegant Definitions: Traditional curricular presentations of the six basic trigonometric functions commonly begin by unnecessarily creating an artificial ("Where did that come from?") ***developmental gap*** in the course. By reducing the common-sensibility of the material, such gaps inhibit mathematical learning and must be minimized.

This paper is primarily about the commonplace "sohcahtoa gap" ("soh: sine = opposite/adjacent", etc.) -- and about one mathematical route for bridging that gap. To a lesser extent, it also is about some ***oscillation-type developmental discontinuities*** – wherein instruction badly vacillates between discordant meanings of some of its rhetoric.

Necessarily, every course follows its instructor's own point-after-point ***mathematical syllabus*** – which might or might not conform to the mathematical syllabus that is woven through some textbook that currently is being used for that course. A mathematical syllabus is ***developmentally continuous*** only to the degree that each of its newly injected "mathematical points" immediately is rationally derived from whatever mathematical theory the students already own.

Developmental discontinuities in mathematical syllabi are major causes for prevalent weaknesses in students' conceptual understanding of curricular mathematics. The "sohcahtoa gap" is by no means the worst among commonplace discontinuities in mathematical syllabi. But that gap nicely illustrates a particular "elegance" kind of instructionally troublesome discontinuities --- and of mathematical bridges for closing such gaps.

The "elegance" gap occurs when the syllabus injects three "sohcahtoa" definitions -- for the "sines", "cosines" and "tangents" of angles – without developing those concepts from within the mathematical thories previously acquired by the students. In an equally naked manner, those soon are followed by labeling of their three reciprocals as "cosecants", "secants", and "cotangents".

The gap comes not from using "sohcahtoa" as a mnemonic device -- which can be a useful tool. Rather, the gap comes from using those three properties as "elegant" formal definitions – instead of conceptually deriving them as useful theorems which follow from ***mathematically derived*** definitions. In their elegance, those definitions largely miss the mathematical essence of the concepts – that the numbers express lengths of circle-parts, relative to circle-radii.-- so, students largely miss it, as well.

Does the curricular choice among alternative, mathematically adequate definitions really matter? Perhaps only to instructors who share a conviction that a major goal for genuine education in mathematics must be student achievement of genuine personal mathematical comprehension of the mathematical theory covered by the course. For others who are concerned only with successfully playing the scholastic game (students' achievement of scores, grades, credits, and credentials), any definition might suffice – as long as it arms the students for successfully running the scholastic gamuts.

Throughout the literature of professional mathematical research, it often is elegant to invoke an important property as a "definition" ... when that same property can otherwise be more tediously developed as a theorem – derived from other definitions that perhaps are far less elegant. The striving for that kind of elegance thrives on authors' expectations that professionals who read the material can accept arbitrary, "out of the blue" definitions as a basis for subsequently digesting whatever mathematical theory then is formally derived from those definitions.

However, such deference to rhetorical elegance often undermines instructional effectiveness. Instead, the caring instructor must choose, as definitions, those properties which provide students with strong conceptual understanding, as needed for functionally internalizing the instructionally targeted mathematical theory.

The typical novice student of trigonometry has not yet achieved the mathematical maturity needed for exploding the "sohcahtoa" definitions into a rational personal theory of the six basic functions. Instead, the thoughtful student easily can be thrown by mathematical jumps ... and might be troubled by such things as, "Why call it that the 'sine' of the angle?" -- and "Why divide?" -- and "Why are we going this route?"

The usual "sohcahtoa" definitions take on forms similar to: "The sine of an angle is the ratio of the length of the opposite side to the length of the hypotenuse." (Wikipedia). But the latter makes sense only for acute angles – not for the other angles of positive or negative rotation that are attended in basic trigonometry. Another stumbling block is its commonplace instructional misuse of the term "ratio". An angle's sine certainly is ***not*** a "ratio" – it is a number – achieved as a quotient.

In careless instructional rhetoric, an oscillation occurs when the term "ratio" sometimes means "quotient" -- and sometimes means "tuple" within an equivalence class of mutually proportional tuples. For students, the proximal vacillation between those two well-separated meanings blurs, hides or destroys the mathematical connections between the two.

Nowhere has the "minor" curricular malady of confusing those two concepts been more disastrous than when teaching "slope" within basic algebra and calculus – in the same mode as used when teaching the tangent function. It has led even to "defining" the slope of a line as a quotient of differences of Cartesian coordinates – which leaves students very much in the dark about why that slope-number is the m(ultiplier) in the mx+b formulas. [The below use of T-square protractors to develop the tan function also can be woven into algebra to enhance the mathematical common-sensibility of the slope-numbers for lines.]

The distinction between number-quotients and tuplic ratios (and the importance of that distinction) becomes quite clear when attending the 3-place tuplic ratios from triangles ... such as 3:4:5 ... as is done in the below development of the six basic functions. When that mathematical distinction is duly attended, the "sohcahtoa" identities and several others become obvious conclusions from more basic geometric definitions.

One improved version of the above quote would be, "The sine of an angle is the quotient of the (positive or negative) height of any point on that angle's terminal ray, along its altitude from the line of that angle's original ray, divided by the length of the hypotenuse out to that point". But even that property – while very useful as a theorem – is instructionally poor as a conceptual "definition" of the sine function. Far more constructive developments are needed and are possible.

Clinical investigations into students' mathematical difficulties with the six basic trigonometric functions lead to more natural and developmentally continuous definitions of those functions – through clarification of the (tuplic) trigonometric ratios. The following development is a result of such clinical research.

Circular Protractors and their Triangles. In elementary plane geometry, an ***angle*** is simply the union of two co-terminal rays. A ***trigon*** is a 3-sided polygon – with the closed cases also being ***triangles***, So, ***trigonometry*** is about using triangles to measure the results from varying some kinds of things. Every simple, closed polygon can be dissected into triangles – and every triangle can be split – in at least one way (maybe 2 or 3) into right triangles. So, much of trigonometry is about right triangles – and about what those say about not-right triangles.

The key is that ***trigonometric angles are*** not merely pairs of coterminal rays; they are ***rotations*** of one (***original-***) ray, into a second (***terminal-***) ray – in the mode of the sweeper on a dial-clock or radar/sonar

screen. So, trigonometric angles are measured in terms of how many ***revolutions*** (***revs***) they express – and ***rotational parts*** of such revolutions.
Every trigonometric angle has its own original and terminal ray. But that same combination of original and terminal rays represents many ***positive and negative*** trigonometric angles. Nonetheless, an angle's original and terminal rays are essential ingredients for defining the six functions.
For purposes of bridging "the sohcahtoa gap", the below development also attends some commonly neglected ingredients: the family of ***full-circle protractors*** that are concentric about the vertex of the angle – and the ***T-square protractors*** that are attached to each of those circles (together called, herein, that angle's ***circle-T protractors***) – and the ***radian measures*** of more protractor parts than just the angles and circular arcs,

A full-circle protractor is a non-degenerate circle with one ray from its center being designated as its original ray – and one direction around its circumference being designated as positive. That circle's T-square protractor consists of the radius along the original ray, and the line tangent to that circle at the outer end of that radius. [The half-way version is the L-square as used in several trades. Most students are more familiar with the semi-circular, half-way versions of circular protractors – but less-so with the use of T-squares or L-squares as protractors.]

[Students also need to know that there is nothing mathematically "special" about trigonometric angles usually increasing counterclockwise. The "navigator's" clockwise-positive orientation comes from northern-hemisphere sun-dials – whose shadows turn "clock" ways – which is why radial clocks do so. The "engineer's" counterclockwise-positive orientation comes from the Sun's earthly passage upward from (earlier) eastward to (later) westward – through "high" noon. Those two orientations merge though the context of trigonometric co-functions. But the trigonometric theory holds for whichever direction of revolution is chosen to be "positive".]
Every trigonometric angle has an infinite family of circular protractors concentric around the endpoint of that angle's original ray – each such circle carrying its own T-square protractor. That angle serves as the ***central angle*** for that system.
[In the case of "unit circles", the T-square also is the "slope-square" that is essential (though usually hidden) within basic algebra.] The six basic functions are geometrically defined in terms of how the revolutions of the central angle's terminal side manifest on its circle-T protractors.
Generalizing Radian-measures; The circular protractor presents arcs from the original ray to all terminal rays – each such arc indicating some part of (positive or negative) a revolution. For purposes of measuring the angles, the circular protractor is factored into parts -- so that the part-measure of that angle is the same, ***regardless of the radius of the protractor***. Circular protractors are thus factored in various convenient ways, according to the needs at hand: quadrants are 1/4 of circles; sextants are 1/6 of circles. octants are 1/8 of circles; day-grees are 1/360 circles.
All trigonometry courses teach and rely on ***radians*** for measuring angles and circular arcs. So, students need to recognize that a radian-arc is that part of a circle whose arc-length is 1-radius (slightly less than one-sixth rev. ... about 0.16 rev. or 57 degrees) – regardless of the size of the circle. But that understanding deepens when ***radian measures are applied also to line segm***ents – a radian being the length of any protractor-part whose length is the same as for that circle's own radii.
[Herein, the term ***radia*** is used to describe line-segment lengths based on radii of associated circles. Although a line-segment might be described as being a number of "radii" in length, the word "radius" connotes a specific kind of parts within the circle – making it inappropriate to speak of lengths of other protractor-parts in terms of "radii". Far better to generalize the widespread use of radian-measures for arc-lengths, so as to include radian-measures also for lengths of line segments. But since "radian", too, already connotes specific parts of a circle, we choose to use "radia", within the context of the circle-T protractors, "radia" for lengths based on the radii of

the respective circles.-- so using it in both the singular and plural meanings.]
Just as the circle's circumference is 2*pi radia, each of its diameters is 2 radia – and every circle is a "unit circle" whose radius is 1 ... exactly 1 radia Thus, we remove a seemingly "special" case: the "unit circle" whose radius is 1. ***All*** circles have radii equal to 1 (radia).

Of greater significance, the use of radia as relative lengths of line segments allow for the six basic trigonometric functions to be developed directly from plane geometry. The "sohcahtoa" and "reciprocal" properties follow as theorems. But the conceptual development is far more meaningful than it is "elegant".

The Angle's Sine-triangles. For each angle, each of its circular protractors generates its own INSIDE right triangle. The latter's hypotenuse (h) is the radius to where the terminal ray intersects the protraction circle. Its altitude-leg (a) is the perpendicular from the line containing the origin-ray, to the outer end of that hypotenuse. The altitude's height is positive, negative, or zero -- depending on the minor arc from the original ray to the terminal ray – and on which direction of rotation has been chosen as "positive".
Likewise, that triangle's base-leg (b) is from the circle's center to the foot of the altitude. When that foot is on the original ray, the base-leg's span is called positive; otherwise it is zero or negative.
Thus, we also remove the seemingly "special" requirement for the trigonometric angle to be "in standard position". Any ray can be used as an original. So is disclosed that there is a practical reason for focusing on angles that are in standard position. Only for their protractors do the bases and altitudes of their inside triangles pair up to give the (base, altitude) Cartesian coordinates for the altitude's endpoint on the circle.
The geometric derivation of the six functions requires geometric meanings for the vocabulary. The word, "sine" simply means "altitude" ... but it immediately presents an angle, a circular protractor about that angle's vertex, and an altitude, from the line of the original ray, to where the terminal ray intersects the circle. One conceptual derivation of that word is as follows.
A partial bow-and-arrow configuration is provided by a rotational angle, one of its circular protractors, and one of the latter's inside triangles. The "hunter's picture" is completed when that angle's negative is appended – along with the latter's own inner triangle. The two opposing altitudes make up one of the circle's chords -- the "bowstring" – the original meaning of "sine". Its "arrow" is along the line of the original ray. Primitive bowstrings were fashioned from ***sinew*** (tendons and such) ... whence comes the word ***sine*** as meaning a chord of a curve, As history would have it, "sine" later became the half-chord of a circular arc – which also is the altitude of an inside right triangle,
So, such inner right triangles are duly called the central angles' ***sine triangles*** within their circular protractors'. Each angle's i family of circular protractors is infinite, and each circle has its own sine triangle for its central angle. All of those triangles lie within the same angular sector of the plane. All of their base-legs are colinear; all of their hypotenuses are colinear; and all of their altitudes are parallel to each other. But even though the triangles are not congruent (because their circles are not) all of their arcs have the same radia-length – as do all of their altitudes, all base-legs and all hypotenuses.
As the trigonometric angle, Θ, continually increases, each sine triangle's altitude (a) increases from a=0 radia to a=1 radia – then decreases to a= ˉ1 radia – then increases back to a=1 radia, The radia-length of the sine-triangle's hypotenuse always is 1 radia ... so the radia-length of the sine does wave between -1 and 1 radia.
So, sine(Θ) is simply the radia-measured altitudes to where the angle's terminal side cuts the rims of its circular protractors. But that definition does not directly reveal how to calculate an angle's sine. So, enter the "soh" theorem.
When a length-scale system is overlaid onto the central angle, the radius of each protraction circle likewise acquires a numeric value, r -- and the sine-triangle's a(ltittude) is sine(Θ)*r. When sine(Θ) already is known

(perhaps by a table, a machine, or a serial formula), the condition, a= sine(Θ)*r, is a handy tool for determining the length of the hypotenuse from that of the altitude – or vice versa.. However, when the central angle's sine is not known, but those two lengths from a sine-triangle are known, those allow for calculating the central angle's sine, as a/h= sine(Θ).

So it is seen that the "soh definition" for the sines of (acute) angles actually presents a way of calculating sines – from lengths of line segments. For "standard position" angles, their sines even can be "soh" calculated directly from coordinates of any point on the terminal ray.

Is a formula that is good for purposes of calculation – and in that context an elegant "definition" – necessarily a good conceptual definition for novice students? Consider this (from Wikipedia). "The slope is defined as the ratio of the "rise" divided by the "run" between two points on a line, " Excepting that "ratio" means "quotient", that property is a great way for calculating the slopes of non-vertical lines in the coordinate plane, when two line-points are known. But that "definition" leaves most students with very little conceptual understanding of what slope-numbers are all about. Much more revealing is the sunburst of lines through the origin – and through the x=1 vertical – in essence, a circle-T protractor.

Just as with line-slopes, the "soh" property mathematically suffices for purposes of calculation – and perhaps as an elegant formal "definition" -- but not as a conceptual definition. There is a major difference between perceiving sine(Θ) as being a numerical quotient of two lengths, and perceiving sine(Θ) as being a directed altitude, measured in radia. The latter interpretation expedites the passage to the sine curve, and also to uses of the sine function to calculate sides of triangles. But more to the point at hand, the altitude-definition of sine naturally derives from the students' own prior knowledge of geometry.

The Angle's Tangent-triangle:s Similar to the circle-T protractor's sine triangles (within the circles) are its outside triangles along its T-squares – called the ***tangent triangles*** for that central angle. Their base-legs are radii from the center out to the tangent lines. Their altitudes – called the angle's ***tangents*** -- are along the tangent line. Those tangents are positive in accord with the direction that is positive for the revolutions, and negative in the opposite direction. All of an angle's tangents have the same radia-measure, regardless of the size of their circles.

Each tangent triangle's hypotenuse lies along the same circle-secant as the angle's terminal ray. Called that angle's ***secants***, those hypotenuses are positive when the terminal ray actually intersects the tangent line, and negative when only the opposing ray intersects the tangent line. (Students are enlightened by the fact that the tangent and secant disappear whenever the terminal ray is parallel to the tangent lines.)

Like the sides of the sine triangles, the sides of the tangent triangles can be measured in radia. When so done, the lengths of an angle's tangent and secant depend only on the angle – not on the radius of the circle.

For each trigonometric angle, and for each of its circle-T protractors, the base-legs of all of its sine and tangent triangles are along the line of that angle's original ray – and all hypotenuses are along the terminal ray. So, all of an angle's sine and tangent triangles (each protraction circle has one of each kind) are similar – regardless of how the lengths are measured.. All of their (altitude: base: hypotenuse) ratios – or ***(a:b:h) ratios*** -- are proportional. In particular, each (tan: radius: sec) ratio is ~ to its (sine: base: radius) ratio.

When an angle's protractors are hidden (as they usually are), each right triangle whose base is from the vertex to a foot on the angle's original ray is both a sine-triangle and a tangent-triangle for an acute base- angle. As a sine-triangle, the hypotenuse is 1 radius of its protraction circle. As a tangent-triangle, the base-leg is 1 radius of a (usually smaller) protraction circle. When those two circles, and the tangent line and secant line for the smaller circle, are overlaid onto that triangle, students can readily see how the given triangle is the sine triangle for the outer circle, and the tangent triangle for the inner circle.

On the other hand, if the right triangle's foot is on the ray opposite to the angle's original ray, the triangle again is a sine triangle, but that triangle is best not regarded as being also a tangent triangle. [Why not? For sure, the T-square that uses that opposite ray does present an "image" tangent line whose triangles identify the tan and sec functions. But in the subsequent passage to the three co-functions, those "offside" image-triangles fail to yield the usual reciprocal properties. Of course, one could accept tangent triangles on both sides, but still use only the original ray's side for developing the co-functions.]

The Angle's Co-triangles For the base-legs of an angle's sine triangles, their distinguishing name comes from the fact that every trigonometric angle has an associated complementary-angle – or ***co-angle*** – for which those two angles add together to give 1/4 rev.

For such purposes, the angle's original ray is rotated 1/4 rev. to provide the original ray for the co-angle – and the two angles share the same terminal ray. But to get the (1/4 rev.) complementation – the co-angle is positively measured in the opposite direction of rotation. As the central angle increases, its co-angle decreases – and vice versa, Thereby, regardless of its number of revolutions, the trigonometric angle and its co-angle add up to 1/4 rev.

A chosen central angle's co-angle has its own sine-triangles and its own tangent triangles – called that angle's ***co-sine*** and ***co-tangent*** triangles. So, the sine, tangent and secant for the chosen angle's co-angle are called the ***co-sine***, ***co-tangent*** and ***co-secant*** for the central angle. When their sides are measured in radia, all of an angle's cosines have the same length, Likewise for all of its cotangents and for all of its cosecants. In its ultimate form, each circle-T protractor thus uses two T-squares – one from the tangent to a radius along the central angle's original ray; and one at the end of the 1/4 rev rotation of that radius.

Within each of its circular protractors, an angle's sine triangle and its cosine triangle constitute a rectangle. One of its corners is the vertex of the central angle, and the opposite corner is on the circle. The shared hypotenuse is one of that rectangle's diagonals.

The central angle's co-sine (altitude) has the same direction and length as the. base-leg of the sine triangle – and the co-sine's base leg has the same direction and length as the. angle's sine. Since those two triangles are congruent, each triangle can be regarded as being that angle's ***(sin, cos, radius) triangle*** or ***(cos, sin, radius) triangle*** – with the latter being in closer harmony with the usual system of Cartesian coordinates. When measured in radia, all radii have length 1.

However, the central angle's (a:b:h) ratio for its sine-triangles is (sine:cos:r) – while that angle's (a:b:h) ratio for its cosine-triangles is (cos:sin:r) – as with 3:4:5 Vs. 4:3:5. That difference manifests as a more prominent difference between the tangent triangle and the cotangent triangle. Both of those latter two are similar to the two (congruent) inner triangles – but the tan and cot triangles typically are far from being congruent to each other.

The Basic Identities Through careful progressive construction of the central; angles' four circle-T triangles, students are easily guided to personally ***derive*** all six traditional "definitions" as personally concluded ***theorems***. In the process, the proportional tuplic ratios for triangles reveal not only how ratios differ from quotients – but also the role of division in converting to, and among unit-ratios.

When the parts of an angle's circle-T protractors are linearly measured in radia, their Pythagorean relationships disclose that $\sin^2 + \cos^2 = 1^2$ – and that $\tan^2 + 1^2 = \sec^2$ – and that $1^2 + \cot^2 = \csc^2$.

In radia, the central angle's own (a:b;h) ratios are (sine:cos:1) – and (tan: 1: sec) – and (1: cot: csc). But, since all three ratios are proportional to each other, each can be converted to each of the others through scalar multiplication.

For (sine:cos:1) –>(tan:1: sec), use 1/cos -- getting [(sin/cos):1:(1/cos)],,,

... disclosing that (sin/cos) = tan -- and that (1/cos) = sec

Handy! It means that tan can be calculated from knowing the ("rise") altitude and the ("run") base – and that sec can be calculated from the base. Useful – but hardly good “definitions”.

For (sine:cos:1) –> (1:cot: csc), use 1/sin -- getting [1:(cos/sin).(1/sin)]
... disclosing that (cos/sin) = cot -- and that (1/sin) = csc.
Likewise handy, but again ....

The four additional conversions provide the other identities.

So it is seen that the usual “reciprocal definitions” – for the three co-functions – actually are properties, which suffice for purposes of calculation. But they largely fail to provide much conceptual understanding about the entailed trigons.

Notably, the tan = sin/cos theorem leads to tan = altitude/ base. On the coordinate plane, it means that the tangent for each non-vertical line’s inclination angle from the horizontal is the dy from any line-point to any other line-point, divided by the (same-directed) dx for those two points. Electronic calculators now have make it possible for students to convert circular measures of angles to their (radia-tangent) slope-numbers for their terminal rays – and vice versa – without calculating. The resulting perception -- that a linear function’s slope-number is a T-square measure of its inclination angle -- goes far toward clarifying what slopes are all about.

Over years of MALEI’s clinical instruction, the preceding development has consistently proven to be much more continuous -- in that students of trigonometry (and above) readily derive the targeted identities from their “middle grade” knowledge of circular protractors and right triangles. Although less frequently, clinical “guided discovery” instruction, has led even some high school algebra-1 students to follow the same path far enough to generate realistic sketches of the sin, cos, and tan curves – and then to use the [sin]. [cos], and [tan] keys on the graphing calculator to calculate lengths of sides on right triangles.