A Comparison of Equality in Computer Algebra and Correctness in Mathematical Pedagogy

Russell Bradford, James Davenport & Chris Sangwin

Universities of Bath, Bath (visiting Waterloo), Birmingham

27 June 2009

Setting

Setting

- A relatively traditional mathematics course, at, say first-year undergraduate level.

Setting

- A relatively traditional mathematics course, at, say first-year undergraduate level.
- **But** Computer-Aided Assessment is in use.

Setting

- A relatively traditional mathematics course, at, say first-year undergraduate level.
- **But** Computer-Aided Assessment is in use.
- One such example is WeBWorK

WeBWorK

(from `http://www.jem-thematic.net/node/586`; our emphasis)

WeBWorK

(from http://www.jem-thematic.net/node/586; our emphasis)

- It's a web-based homework checker.

WeBWorK

(from http://www.jem-thematic.net/node/586; our emphasis)

- It's a web-based homework checker.
- WeBWorK makes homework more effective in calculus and physics courses.

WeBWorK

(from `http://www.jem-thematic.net/node/586`; our emphasis)

- It's a web-based homework checker.
- WeBWorK makes homework more effective in calculus and physics courses.
- The overwhelming majority of students complete all of their homework *correctly* – (sometimes after several attempts).

WeBWorK

(from http://www.jem-thematic.net/node/586; our emphasis)

- It's a web-based homework checker.
- WeBWorK makes homework more effective in calculus and physics courses.
- The overwhelming majority of students complete all of their homework *correctly* – (sometimes after several attempts).
- It is particularly adept at handling mathematics and physics problems.

WeBWorK

(from http://www.jem-thematic.net/node/586; our emphasis)

- It's a web-based homework checker.
- WeBWorK makes homework more effective in calculus and physics courses.
- The overwhelming majority of students complete all of their homework *correctly* – (sometimes after several attempts).
- It is particularly adept at handling mathematics and physics problems.
- The homework is corrected and graded efficiently and *completely*.

Automatically generated problems

Means automatically generated answers *and marking schemes*.

Automatically generated problems

Means automatically generated answers *and marking schemes*.

- Parsing the student's answer (non-trivial — see next slide)

Automatically generated problems

Means automatically generated answers *and marking schemes*.

- Parsing the student's answer (non-trivial — see next slide)
- Is the student's answer mathematically correct?

Automatically generated problems

Means automatically generated answers *and marking schemes*.

- Parsing the student's answer (non-trivial — see next slide)
- Is the student's answer mathematically correct?
- Is the student's answer pedagogically correct?

Automatically generated problems

Means automatically generated answers *and marking schemes*.

- Parsing the student's answer (non-trivial — see next slide)
- Is the student's answer mathematically correct?
- Is the student's answer pedagogically correct?
- So what mark does it get (assuming we are doing more than true/false marking)?

Typical computer aided assessment

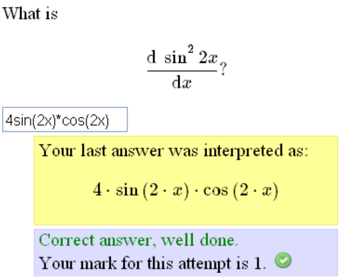

Figure: STACK system [Sangwin2007]

Mathematical Correctness

This is actually a non-trivial problem, even (especially?) with the resources of a computer algebra system behind us.

Mathematical Correctness

This is actually a non-trivial problem, even (especially?) with the resources of a computer algebra system behind us.

- Can we prevent students from regurgitating the question?

Mathematical Correctness

This is actually a non-trivial problem, even (especially?) with the resources of a computer algebra system behind us.

- Can we prevent students from regurgitating the question?
- Can we deal with 'smart alecs'?

Mathematical Correctness

This is actually a non-trivial problem, even (especially?) with the resources of a computer algebra system behind us.

- Can we prevent students from regurgitating the question?
- Can we deal with 'smart alecs'?
- Probably no — a correct answer to "expand $(x+1)^2$" is

$$x^2 + \left(\max_{n \in \mathbf{N}} \exists x, y, z \in \mathbf{N}^* x^n + y^n = z^n \right) x + 1. \qquad (1)$$

Mathematical Correctness

This is actually a non-trivial problem, even (especially?) with the resources of a computer algebra system behind us.

- Can we prevent students from regurgitating the question?
- Can we deal with 'smart alecs'?
- Probably no — a correct answer to "expand $(x+1)^2$" is

$$x^2 + \left(\max_{n \in \mathbf{N}} \exists x, y, z \in \mathbf{N}^* x^n + y^n = z^n\right) x + 1. \tag{1}$$

- How do we define "mathematical correctness"

Mathematical Correctness

This is actually a non-trivial problem, even (especially?) with the resources of a computer algebra system behind us.

- Can we prevent students from regurgitating the question?
- Can we deal with 'smart alecs'?
- Probably no — a correct answer to "expand $(x+1)^2$" is

$$x^2 + \left(\max_{n \in \mathbf{N}} \exists x, y, z \in \mathbf{N}^* x^n + y^n = z^n \right) x + 1. \qquad (1)$$

- How do we define "mathematical correctness"

Mathematical Correctness

This is actually a non-trivial problem, even (especially?) with the resources of a computer algebra system behind us.

- Can we prevent students from regurgitating the question?
- Can we deal with 'smart alecs'?
- Probably no — a correct answer to "expand $(x+1)^2$" is

$$x^2 + \left(\max_{n \in \mathbf{N}} \exists x, y, z \in \mathbf{N}^* x^n + y^n = z^n \right) x + 1. \tag{1}$$

- How do we define "mathematical correctness"

WeBWorK "if it's true at five points it's true"

Mathematical Correctness

This is actually a non-trivial problem, even (especially?) with the resources of a computer algebra system behind us.

- Can we prevent students from regurgitating the question?
- Can we deal with 'smart alecs'?
- Probably no — a correct answer to "expand $(x+1)^2$" is

$$x^2 + \left(\max_{n \in \mathbf{N}} \exists x, y, z \in \mathbf{N}^* x^n + y^n = z^n \right) x + 1. \qquad (1)$$

- How do we define "mathematical correctness"

WeBWorK "if it's true at five points it's true"

STACK Computer algebra — Maxima

Mathematical Correctness

This is actually a non-trivial problem, even (especially?) with the resources of a computer algebra system behind us.

- Can we prevent students from regurgitating the question?
- Can we deal with 'smart alecs'?
- Probably no — a correct answer to "expand $(x+1)^2$" is

$$x^2 + \left(\max_{n\in\mathbf{N}} \exists x, y, z \in \mathbf{N}^* x^n + y^n = z^n\right) x + 1. \qquad (1)$$

- How do we define "mathematical correctness"

WeBWorK "if it's true at five points it's true"

STACK Computer algebra — Maxima

Both have their drawbacks.

Pedagogical Correctness

There's more to this than right/wrong

Pedagogical Correctness

There's more to this than right/wrong

Table: Typical answers: $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$

No.	Student's answer	C.A.	Score
1.	$4\sin 2x\cos 2x$	T	1
2.	$\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$	T	0
3.	$2\sin 2x\cos 2x$	F	0.7
4.	$2\times 2\sin 2x\cos 2x$	T	0.8
5.	$2\sin 4x$	T?	1
6.	$2\sin 2x\cos 2x + 2\sin 2x\cos 2x$	T	0.8
7.	$x/4 - \sin(4*x)/8$	F	0

Pedagogical Correctness

There's more to this than right/wrong

Table: Typical answers: $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$

No.	Student's answer	C.A.	Score
1.	$4\sin 2x \cos 2x$	T	1
2.	$\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$	T	0
3.	$2\sin 2x \cos 2x$	F	0.7
4.	$2 \times 2\sin 2x \cos 2x$	T	0.8
5.	$2\sin 4x$	T?	1
6.	$2\sin 2x \cos 2x + 2\sin 2x \cos 2x$	T	0.8
7.	$x/4 - \sin(4 * x)/8$	F	0

Note that both mathematically "right" and "wrong" answers got 0, and a mathematically "wrong" answer still gets 70%.

Pedagogical Correctness

There's more to this than right/wrong

Table: Typical answers: $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$

No.	Student's answer	C.A.	Score
1.	$4\sin 2x\cos 2x$	T	1
2.	$\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$	T	0
3.	$2\sin 2x\cos 2x$	F	0.7
4.	$2\times 2\sin 2x\cos 2x$	T	0.8
5.	$2\sin 4x$	T?	1
6.	$2\sin 2x\cos 2x + 2\sin 2x\cos 2x$	T	0.8
7.	$x/4 - \sin(4*x)/8$	F	0

Note that both mathematically "right" and "wrong" answers got 0, and a mathematically "wrong" answer still gets 70%. We are **not** dealing with this second problem here!

Marking a wrong answer

Marking a wrong answer

This is not really within the scope of this talk: it falls more in the scope of “buggy rules” [O’Shea, 1982].

Marking a wrong answer

This is not really within the scope of this talk: it falls more in the scope of "buggy rules" [O'Shea, 1982].

- One might ask if it differs from the "true" answer by an additive constant

Marking a wrong answer

This is not really within the scope of this talk: it falls more in the scope of “buggy rules” [O’Shea, 1982].

- One might ask if it differs from the “true” answer by an additive constant
- . . . or a multiplicative constant (more likely for this sort of problem)

Marking a wrong answer

This is not really within the scope of this talk: it falls more in the scope of "buggy rules" [O'Shea, 1982].

- One might ask if it differs from the "true" answer by an additive constant
- . . . or a multiplicative constant (more likely for this sort of problem)
- or . . . .

What is (computer) algebra?

What is (computer) algebra?

The Scratchpad/Axiom characterisation of computer algebra would be that it is working in a "sufficiently rich" order-sorted algebra, i.e. two expressions are **equal** if they are in the same congruence class, for a congruence generated by a "sufficiently rich" set of equations. For $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$ $(\mathrm{D}\sin^2 2x)$, we would have rules like:

What is (computer) algebra?

The Scratchpad/Axiom characterisation of computer algebra would be that it is working in a "sufficiently rich" order-sorted algebra, i.e. two expressions are **equal** if they are in the same congruence class, for a congruence generated by a "sufficiently rich" set of equations. For $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$ $(\mathrm{D}\sin^2 2x)$, we would have rules like:

1. `a*b=b*a;`

What is (computer) algebra?

The Scratchpad/Axiom characterisation of computer algebra would be that it is working in a "sufficiently rich" order-sorted algebra, i.e. two expressions are **equal** if they are in the same congruence class, for a congruence generated by a "sufficiently rich" set of equations. For $\frac{\mathrm{d} \sin^2 2x}{\mathrm{d}x}$ ($\mathrm{D} \sin^2 2x$), we would have rules like:

1. `a*b=b*a`;
2. `m+n`$=m+n$ (e.g. 2+3=5);

What is (computer) algebra?

The Scratchpad/Axiom characterisation of computer algebra would be that it is working in a "sufficiently rich" order-sorted algebra, i.e. two expressions are **equal** if they are in the same congruence class, for a congruence generated by a "sufficiently rich" set of equations. For $\frac{\mathrm{d} \sin^2 2x}{\mathrm{d}x}$ $(\mathrm{D} \sin^2 2x)$, we would have rules like:

1. `a*b=b*a`;
2. `m+n`$= m + n$ (e.g. 2+3=5);
3. `a*c+b*c=(a+b)*c`;

What is (computer) algebra?

The Scratchpad/Axiom characterisation of computer algebra would be that it is working in a "sufficiently rich" order-sorted algebra, i.e. two expressions are **equal** if they are in the same congruence class, for a congruence generated by a "sufficiently rich" set of equations. For $\frac{\mathrm{d} \sin^2 2x}{\mathrm{d}x}$ ($\mathrm{D} \sin^2 2x$), we would have rules like:

1. `a*b=b*a`;
2. `m+n`$=m+n$ (e.g. 2+3=5);
3. `a*c+b*c=(a+b)*c`;
4. `x^ 1=x`;

What is (computer) algebra?

The Scratchpad/Axiom characterisation of computer algebra would be that it is working in a "sufficiently rich" order-sorted algebra, i.e. two expressions are **equal** if they are in the same congruence class, for a congruence generated by a "sufficiently rich" set of equations. For $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$ ($\mathrm{D}\sin^2 2x$), we would have rules like:

1. `a*b=b*a`;
2. `m+n`$=m+n$ (e.g. 2+3$=$5);
3. `a*c+b*c=(a+b)*c`;
4. `x^1=x`;
5. `D(a^n)=n*a^(n-1)*D(a)`;

What is (computer) algebra?

The Scratchpad/Axiom characterisation of computer algebra would be that it is working in a "sufficiently rich" order-sorted algebra, i.e. two expressions are **equal** if they are in the same congruence class, for a congruence generated by a "sufficiently rich" set of equations. For $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$ $(\mathrm{D}\sin^2 2x)$, we would have rules like:

1. `a*b=b*a`;
2. `m+n`$=m+n$ (e.g. 2+3=5);
3. `a*c+b*c=(a+b)*c`;
4. `x^ 1=x`;
5. `D(a^ n)=n*a^ (n-1)*D(a)`;
6. `D(sin(a))=cos(a)*D(a)`.

What is (computer) algebra?

The Scratchpad/Axiom characterisation of computer algebra would be that it is working in a "sufficiently rich" order-sorted algebra, i.e. two expressions are **equal** if they are in the same congruence class, for a congruence generated by a "sufficiently rich" set of equations. For $\frac{\mathrm{d} \sin^2 2x}{\mathrm{d}x}$ $(\mathrm{D} \sin^2 2x)$, we would have rules like:

1. `a*b=b*a`;
2. `m+n`$= m + n$ (e.g. 2+3=5);
3. `a*c+b*c=(a+b)*c`;
4. `x^ 1=x`;
5. `D(a^ n)=n*a^ (n-1)*D(a)`;
6. `D(sin(a))=cos(a)*D(a)`.

What is (computer) algebra?

The Scratchpad/Axiom characterisation of computer algebra would be that it is working in a "sufficiently rich" order-sorted algebra, i.e. two expressions are **equal** if they are in the same congruence class, for a congruence generated by a "sufficiently rich" set of equations. For $\frac{\mathrm{d} \sin^2 2x}{\mathrm{d}x}$ ($\mathrm{D} \sin^2 2x$), we would have rules like:

1. `a*b=b*a`;
2. `m+n`$=m+n$ (e.g. 2+3=5);
3. `a*c+b*c=(a+b)*c`;
4. `x^ 1=x`;
5. `D(a^ n)=n*a^ (n-1)*D(a)`;
6. `D(sin(a))=cos(a)*D(a)`.

Note that it need not be *implemented* this way.

Utility of this model

Utility of this model

In this model, it is indeed true that $\frac{\mathrm{d} \sin^2 2x}{\mathrm{d}x} = 4 \sin 2x \cos 2x$

Utility of this model

In this model, it is indeed true that $\frac{\mathrm{d} \sin^2 2x}{\mathrm{d} x} = 4 \sin 2x \cos 2x$
We can now divide the rules into three categories, which we call *underlying*,

Utility of this model

In this model, it is indeed true that $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x} = 4\sin 2x \cos 2x$
We can now divide the rules into three categories, which we call *underlying*, *venial* (failure to use effectively costs fractions of a point)

Utility of this model

In this model, it is indeed true that $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x} = 4\sin 2x \cos 2x$
We can now divide the rules into three categories, which we call *underlying*, *venial* (failure to use effectively costs fractions of a point) and *fatal* (needing this to get the "right" answer is fatal).

Utility of this model

In this model, it is indeed true that $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x} = 4\sin 2x\cos 2x$
We can now divide the rules into three categories, which we call *underlying*, *venial* (failure to use effectively costs fractions of a point) and *fatal* (needing this to get the "right" answer is fatal).
Our equations are then categorised as follows.

U1 `a*b=b*a;`

Utility of this model

In this model, it is indeed true that $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x} = 4\sin 2x\cos 2x$
We can now divide the rules into three categories, which we call *underlying*, *venial* (failure to use effectively costs fractions of a point) and *fatal* (needing this to get the "right" answer is fatal). Our equations are then categorised as follows.

U1 `a*b=b*a`;

V2 `m+n`$=m+n$ (e.g. 2+3=5) (and $-$);

Utility of this model

In this model, it is indeed true that $\frac{\mathrm{d} \sin^2 2x}{\mathrm{d}x} = 4 \sin 2x \cos 2x$
We can now divide the rules into three categories, which we call *underlying*, *venial* (failure to use effectively costs fractions of a point) and *fatal* (needing this to get the "right" answer is fatal). Our equations are then categorised as follows.

- U1 `a*b=b*a`;
- V2 `m+n`$=m+n$ (e.g. 2+3=5) (and $-$);
- V3 `a*c+b*c=(a+b)*c`;

Utility of this model

In this model, it is indeed true that $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x} = 4\sin 2x\cos 2x$
We can now divide the rules into three categories, which we call *underlying*, *venial* (failure to use effectively costs fractions of a point) and *fatal* (needing this to get the "right" answer is fatal). Our equations are then categorised as follows.

- U1 `a*b=b*a;`
- V2 `m+n`$=m+n$ (e.g. 2+3=5) (and $-$);
- V3 `a*c+b*c=(a+b)*c;`
- V4 `x^ 1=x;`

Utility of this model

In this model, it is indeed true that $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x} = 4\sin 2x\cos 2x$
We can now divide the rules into three categories, which we call *underlying*, *venial* (failure to use effectively costs fractions of a point) and *fatal* (needing this to get the "right" answer is fatal). Our equations are then categorised as follows.

- U1 `a*b=b*a;`
- V2 `m+n`$=m+n$ (e.g. 2+3=5) (and $-$);
- V3 `a*c+b*c=(a+b)*c;`
- V4 `x^ 1=x;`
- F1 `D(a^ n)=n*a^ (n-1)*D(a);`

Utility of this model

In this model, it is indeed true that $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x} = 4\sin 2x\cos 2x$
We can now divide the rules into three categories, which we call *underlying*, *venial* (failure to use effectively costs fractions of a point) and *fatal* (needing this to get the "right" answer is fatal). Our equations are then categorised as follows.

- U1 `a*b=b*a`;
- V2 `m+n`$=m+n$ (e.g. 2+3=5) (and $-$);
- V3 `a*c+b*c=(a+b)*c`;
- V4 `x^ 1=x`;
- F1 `D(a^ n)=n*a^ (n-1)*D(a)`;
- F2 `D(sin(a))=cos(a)*D(a)`.

Answers analysed

Let $\mathcal{U}$ be the set of underlying equations, $\mathcal{V}$ be the underlying and venial ones, and $\mathcal{F}$ the set of them all.

Answers analysed

Let $\mathcal{U}$ be the set of underlying equations, $\mathcal{V}$ be the underlying and venial ones, and $\mathcal{F}$ the set of them all.

Table: Analysed answers: $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x} = 4\cos 2x \sin 2x$

No.	Student's answer	relation	Score
1.	$4\sin 2x \cos 2x$	$\equiv_{\mathcal{U}}$	1
2.	$\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$	$\equiv_{\mathcal{F}}$	0
3.	$2\sin 2x \cos 2x$	F	buggy
4.	$2 \times 2\sin 2x \cos 2x$	$\equiv_{\mathcal{V}}$	0.8
5.	$2\sin 4x$	none	0
6.	$2\sin 2x \cos 2x + 2\sin 2x \cos 2x$	$\equiv_{\mathcal{V}}$	0.8
7.	$x/4 - \sin(4 * x)/8$	none	0

Answers analysed

Let $\mathcal{U}$ be the set of underlying equations, $\mathcal{V}$ be the underlying and venial ones, and $\mathcal{F}$ the set of them all.

Table: Analysed answers: $\frac{\mathrm{d} \sin^2 2x}{\mathrm{d}x} = 4\cos 2x \sin 2x$

No.	Student's answer	relation	Score
1.	$4\sin 2x \cos 2x$	$\equiv_{\mathcal{U}}$	1
2.	$\frac{\mathrm{d} \sin^2 2x}{\mathrm{d}x}$	$\equiv_{\mathcal{F}}$	0
3.	$2\sin 2x \cos 2x$	F	buggy
4.	$2 \times 2\sin 2x \cos 2x$	$\equiv_{\mathcal{V}}$	0.8
5.	$2\sin 4x$	none	0
6.	$2\sin 2x \cos 2x + 2\sin 2x \cos 2x$	$\equiv_{\mathcal{V}}$	0.8
7.	$x/4 - \sin(4 * x)/8$	none	0

Note that answer 5 is marked wrong, since trigonometric contraction is not one of our rules. It probably should be, but we need a digression.

What is the “right” answer

Assuming we do not know about trigonometric contraction, most people would say $4 \sin 2x \cos 2x$ (or $4 \cos 2x \sin 2x$, which is equivalent under $\mathcal{U}$).

What is the “right” answer

Assuming we do not know about trigonometric contraction, most people would say $4\sin 2x\cos 2x$ (or $4\cos 2x\sin 2x$, which is equivalent under $\mathcal{U}$). But mathematically this is

$$\sin 2x\cos 2x + 3\cos 2x\sin 2x$$

(and many other expressions).

What is the "right" answer

Assuming we do not know about trigonometric contraction, most people would say $4\sin 2x\cos 2x$ (or $4\cos 2x\sin 2x$, which is equivalent under $\mathcal{U}$). But mathematically this is

$$\sin 2x\cos 2x + 3\cos 2x\sin 2x$$

(and many other expressions). Of course, we really want the "simplest" answer. What does "simplify" mean?

What is the "right" answer

Assuming we do not know about trigonometric contraction, most people would say $4 \sin 2x \cos 2x$ (or $4 \cos 2x \sin 2x$, which is equivalent under $\mathcal{U}$). But mathematically this is

$$\sin 2x \cos 2x + 3 \cos 2x \sin 2x$$

(and many other expressions). Of course, we really want the "simplest" answer. What does "simplify" mean?

- "simplify by removing brackets and collecting like terms".

What is the "right" answer

Assuming we do not know about trigonometric contraction, most people would say $4 \sin 2x \cos 2x$ (or $4 \cos 2x \sin 2x$, which is equivalent under $\mathcal{U}$). But mathematically this is

$$\sin 2x \cos 2x + 3 \cos 2x \sin 2x$$

(and many other expressions). Of course, we really want the "simplest" answer. What does "simplify" mean?

- "simplify by removing brackets and collecting like terms".
- "factor and cancel like terms" (in the same book!)

What is the “right” answer

Assuming we do not know about trigonometric contraction, most people would say $4 \sin 2x \cos 2x$ (or $4 \cos 2x \sin 2x$, which is equivalent under $\mathcal{U}$). But mathematically this is

$$\sin 2x \cos 2x + 3 \cos 2x \sin 2x$$

(and many other expressions). Of course, we really want the “simplest” answer. What does “simplify” mean?

- “simplify by removing brackets and collecting like terms”.
- “factor and cancel like terms” (in the same book!)
- “do what I’ve just shown you” (often).

What is the "right" answer

Assuming we do not know about trigonometric contraction, most people would say $4 \sin 2x \cos 2x$ (or $4 \cos 2x \sin 2x$, which is equivalent under $\mathcal{U}$). But mathematically this is

$$\sin 2x \cos 2x + 3 \cos 2x \sin 2x$$

(and many other expressions). Of course, we really want the "simplest" answer. What does "simplify" mean?

- "simplify by removing brackets and collecting like terms".
- "factor and cancel like terms" (in the same book!)
- "do what I've just shown you" (often).
- "Give me the answer I want" (Classes préparatoires professors to JHD).

What is the "right" answer

Assuming we do not know about trigonometric contraction, most people would say $4\sin 2x\cos 2x$ (or $4\cos 2x\sin 2x$, which is equivalent under $\mathcal{U}$). But mathematically this is

$$\sin 2x\cos 2x + 3\cos 2x\sin 2x$$

(and many other expressions). Of course, we really want the "simplest" answer. What does "simplify" mean?

- "simplify by removing brackets and collecting like terms".
- "factor and cancel like terms" (in the same book!)
- "do what I've just shown you" (often).
- "Give me the answer I want" (Classes préparatoires professors to JHD).
- [Carette 2004] "The/A shortest equivalent expression".

What is the "right" answer

Assuming we do not know about trigonometric contraction, most people would say $4\sin 2x\cos 2x$ (or $4\cos 2x\sin 2x$, which is equivalent under $\mathcal{U}$). But mathematically this is

$$\sin 2x\cos 2x + 3\cos 2x\sin 2x$$

(and many other expressions). Of course, we really want the "simplest" answer. What does "simplify" mean?

- "simplify by removing brackets and collecting like terms".
- "factor and cancel like terms" (in the same book!)
- "do what I've just shown you" (often).
- "Give me the answer I want" (Classes préparatoires professors to JHD).
- [Carette 2004] "The/A shortest equivalent expression".

What is the "right" answer

Assuming we do not know about trigonometric contraction, most people would say $4 \sin 2x \cos 2x$ (or $4 \cos 2x \sin 2x$, which is equivalent under $\mathcal{U}$). But mathematically this is

$$\sin 2x \cos 2x + 3 \cos 2x \sin 2x$$

(and many other expressions). Of course, we really want the "simplest" answer. What does "simplify" mean?

- "simplify by removing brackets and collecting like terms".
- "factor and cancel like terms" (in the same book!)
- "do what I've just shown you" (often).
- "Give me the answer I want" (Classes préparatoires professors to JHD).
- [Carette 2004] "The/A shortest equivalent expression".

"The right answer" is "*a shortest expression under* $\equiv_{\mathcal{F}}$".

What is the "right" answer

Assuming we do not know about trigonometric contraction, most people would say $4\sin 2x \cos 2x$ (or $4\cos 2x \sin 2x$, which is equivalent under $\mathcal{U}$). But mathematically this is

$$\sin 2x \cos 2x + 3\cos 2x \sin 2x$$

(and many other expressions). Of course, we really want the "simplest" answer. What does "simplify" mean?

- "simplify by removing brackets and collecting like terms".
- "factor and cancel like terms" (in the same book!)
- "do what I've just shown you" (often).
- "Give me the answer I want" (Classes préparatoires professors to JHD).
- [Carette 2004] "The/A shortest equivalent expression".

"The right answer" is "*a shortest expression under* $\equiv_{\mathcal{F}}$". It had better be the case that only $\mathcal{U}$ can produce equivalent expressions of the same length.

Answers re-analysed

Add various fraction-simplifying rules to $\mathcal{V}$, and

U2 `sin a * cos a =` $\frac{1}{2}$`sin 2*a`.

Answers re-analysed

Add various fraction-simplifying rules to $\mathcal{V}$, and

U2 `sin a * cos a =` $\frac{1}{2}$`sin 2*a`.

Answers re-analysed

Add various fraction-simplifying rules to $\mathcal{V}$, and

U2 `sin a * cos a =` $\frac{1}{2}$`sin 2*a`.

Table: Re-analysed answers: $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x} = 2\sin 4x$

No.	Student's answer	relation	Score
1.	$4\sin 2x \cos 2x$	$\equiv_{\mathcal{U}}$	1
2.	$\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$	$\equiv_{\mathcal{F}}$	0
3.	$2\sin 2x \cos 2x$	F	buggy
4.	$2 \times 2\sin 2x \cos 2x$	$\equiv_{\mathcal{V}}$	0.8
5.	$2\sin 4x$	$=$	1
6.	$2\sin 2x \cos 2x + 2\sin 2x \cos 2x$	$\equiv_{\mathcal{V}}$	0.8
7.	$x/4 - \sin(4 * x)/8$	none	0

Answers re-analysed

Add various fraction-simplifying rules to $\mathcal{V}$, and

U2 `sin a * cos a =` $\frac{1}{2}$`sin 2*a`.

Table: Re-analysed answers: $\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x} = 2\sin 4x$

No.	Student's answer	relation	Score
1.	$4\sin 2x \cos 2x$	$\equiv_{\mathcal{U}}$	1
2.	$\frac{\mathrm{d}\sin^2 2x}{\mathrm{d}x}$	$\equiv_{\mathcal{F}}$	0
3.	$2\sin 2x \cos 2x$	F	buggy
4.	$2 \times 2\sin 2x \cos 2x$	$\equiv_{\mathcal{V}}$	0.8
5.	$2\sin 4x$	$=$	1
6.	$2\sin 2x \cos 2x + 2\sin 2x \cos 2x$	$\equiv_{\mathcal{V}}$	0.8
7.	$x/4 - \sin(4 * x)/8$	none	0

Note that answer 5 is now precisely right.

Variations on a theme

- We could have added rule U2 to the set $\mathcal{V}$, rather than to $\mathcal{U}$. This would then mean that $2 \sin 4x$ was now right, but $4 \sin 2x \cos 2x$, although still mathematically correct, only scores 0.8, since it is only equivalent to the right answer under venial rules.

Variations on a theme

- We could have added rule U2 to the set $\mathcal{V}$, rather than to $\mathcal{U}$. This would then mean that $2 \sin 4x$ was now right, but $4 \sin 2x \cos 2x$, although still mathematically correct, only scores 0.8, since it is only equivalent to the right answer under venial rules.
- We could create a new class $\mathcal{V}'$ of "weakly venial" rules between $\mathcal{U}$ and $\mathcal{V}$, with U2 in it, and say that $\equiv_{\mathcal{V}'}$ was worth, say, 0.9 rather than 0.8.

Variations on a theme

- We could have added rule U2 to the set $\mathcal{V}$, rather than to $\mathcal{U}$. This would then mean that $2\sin 4x$ was now right, but $4\sin 2x\cos 2x$, although still mathematically correct, only scores 0.8, since it is only equivalent to the right answer under venial rules.
- We could create a new class $\mathcal{V}'$ of "weakly venial" rules between $\mathcal{U}$ and $\mathcal{V}$, with U2 in it, and say that $\equiv_{\mathcal{V}'}$ was worth, say, 0.9 rather than 0.8.
- The teacher could vary the approach over time, saying "from now on, I expect you to use trigonometric contraction where appropriate", and move U2 from $\mathcal{U}$ to $\mathcal{V}'$, and maybe on to $\mathcal{V}$ after a couple of weeks.

Variations on a theme

- We could have added rule U2 to the set $\mathcal{V}$, rather than to $\mathcal{U}$. This would then mean that $2\sin 4x$ was now right, but $4\sin 2x\cos 2x$, although still mathematically correct, only scores 0.8, since it is only equivalent to the right answer under venial rules.
- We could create a new class $\mathcal{V}'$ of "weakly venial" rules between $\mathcal{U}$ and $\mathcal{V}$, with U2 in it, and say that $\equiv_{\mathcal{V}'}$ was worth, say, 0.9 rather than 0.8.
- The teacher could vary the approach over time, saying "from now on, I expect you to use trigonometric contraction where appropriate", and move U2 from $\mathcal{U}$ to $\mathcal{V}'$, and maybe on to $\mathcal{V}$ after a couple of weeks.
- Indeed, one could imagine a stronger form of $\mathcal{V}$, which cost 50% of the marks.

Conclusions

Conclusions

- We can't let *any* algebra get at the student's input before we do!

Conclusions

- We can't let *any* algebra get at the student's input before we do!
- This is going to be even more important as we develop tests like 'factor', 'express as partial fractions' etc.

Conclusions

- We can't let *any* algebra get at the student's input before we do!
- This is going to be even more important as we develop tests like 'factor', 'express as partial fractions' etc.
- This formalism may actually help a teacher *explain why*, rather than just say "I expected you to".

Conclusions

- We can't let *any* algebra get at the student's input before we do!
- This is going to be even more important as we develop tests like 'factor', 'express as partial fractions' etc.
- This formalism may actually help a teacher *explain why*, rather than just say "I expected you to".
- We do *not* preclude use of the full power of a computer algebra system — "the system thinks your answer is right, but you'd better get it marked manually".