Notes on a (very) Elementary Set Theory—Part III

1 Induction with respect to a general relation

We are in informal mathematics until otherwise stated.[1]

Suppose that we have a set X that we will keep for a while as our "relative universe".[2] This means, in particular, that object variables of formulas and terms vary over (i.e., take values from) X, and also that whenever we write $(\forall x)A$ or $(\exists x)A$ we really mean $(\forall x \in X)A$ or $(\exists x \in X)A$ respectively.

Moreover, let us fix attention on a binary relation $\prec$ that is defined on X. That is, $\prec$ (like $=, \subseteq, \subset, \in, <$ and many others) accepts just two inputs t and s (arbitrary terms) and leads to the formula $t \prec s$. This formula, depending on the values of t and s may evaluate to true or false (**t** or **f**). We may read it as "t precedes s" or "t is before s" or "t is lower than s".

There is no special significance in the shape of $\prec$. This symbol can stand for any relation whatsoever, but it will have one important property. This "important property" will entail that the relation "behaves like" $<, \subset$ in that it will turn out that $x \prec x$ will be always false. This is why there is no extra line below the symbol (unlike $\subseteq, \leq$ and $\preceq$).

Definition 1.1. We say that $\prec$ satisfies the *inductiveness condition*, in short IC, iff the following is true for any formula A:

$$(\forall x)\big((\forall z \prec x)A[z] \Rightarrow A[x]\big) \Rightarrow (\forall x)A[x] \tag{1}$$

■

(1) above looks exactly like the CVI schema, the only difference being the use of "$\prec$" in (1) rather than the specific "$<$" over the specific set $\mathbb{N}$.

Just as with CVI, we can prove formulas $A[x]$ where x varies over X by a CVI-like induction. Let us call it GCVI for "generalised CVI". The protocol is:

---

[1] But we still apply proper logic to get results proved. In particular, we are responsible for what we assume at every step. Our "assumptions" must be realistic and not wishful thinking.

[2] Recall that the absolute universe, $\mathbb{V}$, is not a set.

GCVI 1. (I.H.) Assume $(\forall z \prec x)A[z]$, or, in words, assume that $A[z]$ is true for all $z \prec x$.
GCVI 2. (Goto) Prove $A[x]$.
GCVI 3. (Conclusion) $(\forall x)A[x]$ has been proved.

Remember what we must understand as long as we work in the "universe" X: The conclusion part is shorthand for "$(\forall x \in X)A[x]$ has been proved".

OK, but how can we *easily* tell whether some relation $\prec$ on some set X satisfies IC, and that therefore we can use GCVI with respect to it? Let us work towards answering this.

Definition 1.2. We say that $\prec$ satisfies the *minimality condition*, in short MC, iff the following is true for any formula A:

$$(\exists x)A[x] \Rightarrow (\exists x)\big(A[x] \land \neg(\exists z \prec x)A[z]\big) \tag{2}$$

■

First we note that (1) in 1.1 and (2) in 1.2 are logically equivalent. Indeed,

$$\begin{aligned}
&(\forall x)\big((\forall z \prec x)A[z] \Rightarrow A[x]\big) \Rightarrow (\forall x)A[x]\\
\equiv&\\
&\neg(\forall x)A[x] \Rightarrow \neg(\forall x)\big((\forall z \prec x)A[z] \Rightarrow A[x]\big)\\
\equiv&\\
&(\exists x)\neg A[x] \Rightarrow (\exists x)\big((\forall z \prec x)A[z] \land \neg A[x]\big)\\
\equiv&\\
&(\exists x)\neg A[x] \Rightarrow (\exists x)\big(\neg A[x] \land \neg(\exists z \prec x)\neg A[z]\big)
\end{aligned}$$

This calculation shows that I get (1) (top line) if I have (2) [(2) is a schema, and I can use $\neg A$ everywhere in (2) instead of A—but doing so gives me the bottom line]. Replacing all A's in the calculation by $\neg A$ we get the converse: If I have (1) I can get (2).

OK, now we are almost there (theorem 1.4 below). But let us digress first with an application of the equivalence of IC and MC in the case where $X = \mathbb{N}$ and $\prec$ is $<$.

Example 1.3. We know from our Peano arithmetic notes and from class that $<$ on $\mathbb{N}$ satisfies IC ("admits induction" as GS put it). Indeed we have proved the CVI schema in this case formally, strictly from the Peano axioms.

But then $<$ satisfies MC as well by the equivalence between IC and MC.

The ancient Greeks used proofs by MC on $\mathbb{N}$, nowadays called proofs by the "least principle" (see remark on this nomenclature on p.11). Such proofs are equivalent to induction proofs and they prove statements like "$(\forall x)A[x]$".

They go like this (informally):

(i) By way of contradiction, assume $\neg(\forall x)A[x]$, that is, assume $(\exists x)\neg A[x]$.
(ii) By MC of $<$ on $\mathbb{N}$ there is an $n \in \mathbb{N}$ such that $\neg A[n]$ is true, but for no $m < n$ is $\neg A[m]$ true. I.e., n is *least* for which $\neg A[n]$.
(iii) Now take advantage of what $A[x]$ actually is, and of whatever theorems you know, to get a contradiction.

Let us apply the recipe (i)–(iii) to an ancient (and famous) theorem: *$\sqrt{2}$ is "irrational", i.e., there are no positive integers m and n such that $\sqrt{2} = m/n$.* Another way of putting it is $(\forall m > 0)(\forall n > 0)(\sqrt{2} \neq m/n)$.

Well, assume the contrary, that such m, n exist. Of these choose the pair with *smallest* n. Thus we have assumed

$$\sqrt{2} = \frac{m}{n} \tag{1}$$

with n being the smallest that works. Now square (1) to get

$$2n^2 = m^2 \tag{2}$$

Note that m cannot be odd, $m = 2k+1$, for then $m^2 = 4k^2 + 4k + 1$ which is also odd—so it cannot be even as well, as (2) indicates.

So $m = 2k$. Eliminate m in (2): We get

$$n^2 = 2k^2 \tag{3}$$

and working exactly as before we conclude that n is even, say, $n = 2r$. By (1) we get

$$\sqrt{2} = \frac{m}{n} = \frac{k}{r}$$

and we have expressed $\sqrt{2}$ with a denominator $r < n$ contradicting that n was the smallest that worked. Done! ■

Theorem 1.4. *$\prec$ on X has MC (and therefore also IC) iff there is no infinite left-chain*

$$\ldots a''' \prec a'' \prec a' \prec a \tag{1}$$

of elements of X.

Proof. (Informal)

(I) I assume that $\prec$ on X has MC—i.e., I have (2) in 1.2. I prove that I cannot have infinite chains like (1).
Well, by way of contradiction assume that I do have a chain like (1). Let S be the set of all the members in the chain.

Wait a minute! How do I know that S is a set? What if it is too big?
Well, it can't be, since all members of S are in X, and thus we can use the axiom of subsets to see that S is therefore a set.
Why is it important that the collection S be a set? Because below I am going to use the formula "$x \in S$". This will be well-formed, i.e., syntactically correct, only if the constant[3] S and variable x have the correct types (atom or set).

Then, by (2)

$$(\exists x)x \in S \Rightarrow (\exists x)\big(x \in S \wedge \neg(\exists z \prec x)z \in S\big)$$

from which, since $(\exists x)x \in S$ says "$S \neq \emptyset$" and is therefore true, I get

$$(\exists x)\big(x \in S \wedge \neg(\exists z \prec x)z \in S\big)$$

This says that some member of S has no $\prec$-predecessor which is false by construction of S. This contradiction proves that I cannot have infinite chains if I have (2).

(II) Conversely, I assume that I cannot have infinite chains like (1), and I prove that $\prec$ then must have MC.
To prove (2) of 1.2 assume the left hand side, i.e., let $(\exists x)A[x]$. Thus, for some w, $A[w]$ is true.[4]

---

[3] S names a specific set.

[4] Formally that would be done by assuming $A[w]$, where w is "fresh". Informally we say "let w be a value that makes A true", a legitimate "let" since we are told that such values exist!

I will show that there is a y such that $A[y]$ is true, and moreover, that for no $z \prec y$ is $A[z]$ true—in symbols, I'll show $(\exists y)(A[y] \wedge \neg(\exists z \prec y)A[z])$, as I must, towards (2).
Well, consider the following process:

L1. If there is no $z \prec w$ such that $A[z]$ then we are lucky! Take $y := w$. Done!

L2. Otherwise, there is a $w' \prec w$ such that $A[w']$.
If there is no $z \prec w'$ such that $A[z]$ then we are lucky in our 2nd try. Take $y := w'$. Done!

L3. Otherwise, there is a $w'' \prec w'$ such that $A[w'']$.
If there is no $z \prec w''$ such that $A[z]$ then we are lucky in our 3rd try. Take $y := w''$. Done!

L4. Otherwise, there is a $w''' \prec w''$ such that $A[w''']$.
If there is no $z \prec w'''$ such that $A[z]$ then we are lucky in our 4th try. Take $y := w'''$. Done!

L5. And so on.

Now the above process builds a chain $\ldots w''' \prec w'' \prec w' \prec w$ in X. By assumption, this chain must terminate, and hence so must the process **L1, L2, L3,** .... But wherever this process terminates it does so because we found a y that works. ■

Definition 1.5. Given a formula $A[x]$. We say that an element a of X such that $A[a]$ is true, but for no $b \prec a$ is $A[b]$ true, is *a minimal element for* $A[x]$.

In the special case where $A[x]$ is the formula $x \in T$ where $\emptyset \neq T \subseteq X$, we simply say that "$a$ is a minimal element of T". ■

Thus, the process **L1, L2, L3,** ... above constructs a minimal element y for $A[x]$.

Definition 1.6. We say that a $\prec$ on X is *well-founded* iff there is no infinite descending chain $\ldots a'' \prec a' \prec a$ of elements $a, a', a'', \ldots$ in X. ■

Our results have shown that a $\prec$ is well-founded iff it has MC iff it has IC ("admits induction").

If $\prec$ is well-founded then it is *irreflexive*, that is, $(\forall x)\neg x \prec x$. Indeed, if not, we have $a \prec a$ for some $a \in X$, and that leads to an infinite chain $\ldots a \prec a \prec a \prec a$.

(Do you remember what “” means? If not check Part I before continuing!)

By the axiom of foundation, $\in$—a relation on “everything” (i.e., on $\mathbb{V}$)—is well-founded in the sense of 1.6. Thus, for any A we have $\in$-MC:

$$(\exists x)A[x] \Rightarrow (\exists x)\big(A[x] \wedge \neg(\exists z \in x)A[z]\big) \tag{1}$$

and can therefore also prove “properties” (i.e., formulas $A[x]$) of arbitrary sets “x” by GCVI with respect to $\in$ using

$$(\forall x)\big((\forall z \in x)A[z] \Rightarrow A[x]\big) \Rightarrow (\forall x)A[x]$$

A *special case* of (1) is when we take $A[x]$ to be the formula $x \in y$. Then (1) becomes

$$(\exists x)x \in y \Rightarrow (\exists x)\big(x \in y \wedge \neg(\exists z \in x)z \in y\big)$$

or, since it does not matter what set y I have in mind, I can generalise

$$(\forall y)\Big((\exists x)x \in y \Rightarrow (\exists x)\big(x \in y \wedge \neg(\exists z \in x)z \in y\big)\Big) \tag{2}$$

which is—you will recall that I so suggested in Part II—the formal version of the axiom of foundation (see -remark that follows axiom 2.1).

Let us see why in some detail. Now, in Part II I claimed that (2) is *equivalent* to the absence of infinite descending chains such as

$$\ldots a_3 \in a_2 \in a_1 \in a_0 \tag{3}$$

However, here, so far, I have only argued one direction: That if (3) is impossible, then I must have (2)—just because this is a *special* case of (1), which I have by MC $\equiv$ IC $\equiv$ foundation. But, what about the converse: if I have the *special* case (2), does the general case, (1), follow—or, equivalently, can I prove the impossibility of (3) by assuming (2)?

Yes, I can.

Assume that we know that (2) holds. I will prove the impossibility of (3).

By contradiction, suppose that I do have (3) for some $a_0, a_1, a_2, \ldots$

Collect all the $a_0, a_1, a_2, a_3, \ldots$ of the infinite chain.

Main claim. *This collection is a set.* We will denote it by S.

To prove the claim, note that the collection S is not any bigger than the set of natural numbers $\mathbb{N}$, therefore it is a set too.

Indeed, to see why S is not bigger, imagine that all the members of $\mathbb{N}$ and all the members of S were invited to a dance of pairs. I am sure they would accept, because there is a way to form pairs and use all the members of each set without having anyone from either set dancing alone: Just pair 0 with a_0 and, in general, pair n with a_n.

This concludes the (informal) proof of the **Main claim**.

I can now contradict the boxed assumption that I took earlier on, concluding that (3) is impossible.

Well, now that I know that S is a set, specialising[5] (2) I get

$$(\exists x)x \in S \Rightarrow (\exists x)\big(x \in S \wedge \neg(\exists z \in x)z \in S\big) \tag{4}$$

By construction of S, the lhs of $\Rightarrow$ in (4) is true (just as we argued in theorem 1.4). Then so is the rhs.

But the rhs is also false, giving us a contradiction: Indeed, the rhs says that for some "special" x in S there is no z, also in S, such that $z \in x$. But this is incorrect! This x has to have the form a_n for an appropriate $n \geq 0$. But then, there is a $z \in S$ that is also in x; namely, $z = a_{n+1}$ will do.

Several observations are now in order:

(A) The proof of the claim used a valid principle of comparing sizes of two collections in order to conclude that one of them is a set. This principle has a formalisation that is called the *axiom of replacement.* I chose not to introduce all the axioms of set theory in order to keep matters "semi-formal" and easy.[6] The axiom of replacement says that *if each member of a set T is replaced by a (possibly different) set or atom, then the resulting collection is also a set.* Note how S above is obtained from the set $\mathbb{N}$ by replacing n by a_n for $n \geq 0$.
One of the reasons I did not introduce this "easy sounding" axiom formally is that its use in a formal setting—e.g., to formalise the proof

[5] The variable y in (2) is a set variable. It was crucial to know that S is a set in order to do the substitution $[y := S]$.

[6] The omitted axioms are those for *replacement*, *infinity* and *choice*. Moreover I have allowed us to intuitively accept that atoms have no elements, but strictly speaking one ought to formalise this via an axiom that for all variables x of type atom says $(\forall x)(\forall y)y \notin x$.

of the main claim above—often requires yet another axiom, the "axiom of infinity", which, essentially, says that we can build (or implement) a perfect copy of $\mathbb{N}$ within set theory. You see, in a totally formal setting we are not allowed to pretend that we know anything about $\mathbb{N}$. The axioms of set theory do not speak of $\mathbb{N}$, but it turns out that they have enough power to allow us to build a replica of $\mathbb{N}$. But to do so is a very long story that we are not going to get into.

(B) In the proof of MC $\equiv$ foundation for an arbitrary $\prec$ on a set X (see 1.4) *we did not need the axiom of replacement* in order to argue that the collection of members of the chain (1) (p.4) is a set. What we used was our assumption that $\prec$ is a relation on X, which allowed us to use the axiom of subsets. Here however the specific (concrete) $\prec$ is $\in$, a relation on the collection $\mathbb{V}$ that is not a set and we needed the overkill.

(C) Now that I have mentioned (B) above, I must draw your attention to the earlier sentence "However, here, so far, I have only argued that if (3) is impossible, then I must have (2)—just because this is a *special* case of (1), which I have by MC $\equiv$ IC $\equiv$ foundation." This assertion contains a small white lie: The equivalence MC $\equiv$ IC $\equiv$ foundation was proved in 1.4 under the strict assumption that $\prec$ acts on a set X. Do I still have the equivalence if $\prec$ acts on a non-set collection, as is the situation with $\in$?
Yes. First off, the equivalence MC $\equiv$ IC is independent of where $\prec$ acts; indeed it is a logical equivalence. As for the MC $\equiv$ foundation part of the equivalence, note that direction (II) in the proof of 1.4 is, again, independent of where $\prec$ acts. It is direction (I) that we proved relying on the assumption that the collection on which $\prec$ acts—i.e., X—is a set. This assumption is not crucial and can be avoided at the expense of a more sophisticated proof and the introduction of more axioms: Indeed, we can still prove that the S constructed in the proof of 1.4 is a set, exactly as we did for the present S above, by using the axiom of replacement informally. This is because the S in the proof of 1.4 is obtained from $\mathbb{N}$ by the replacements $0 \mapsto a$, $1 \mapsto a'$, $2 \mapsto a''$, $3 \mapsto a'''$, and so on in the obvious pattern.

Since $\in$—a special case of our abstract $\prec$—has MC, IC and foundation, but is not transitive, we must be careful to recognise that "$\prec$*, in general, is not transitive*".

It turns out that it is relatively easy to verify that $\prec$ is well-founded in many cases of interest (next section). This also verifies that we can prove theorems about the members of the "universe" X by $\prec$-induction (GCVI).

2 Concrete cases of "$\prec$". Some GCVI proofs

Example 2.1. Our first example uses as X the set of all propositional (Boolean) formulas and $\prec$ is the *proper subformula* relation.

First off, recall that formulas are defined inductively on the alphabet

$$\{p, q, r, p', q', r', \ldots; \mathit{false}; \mathit{true}; (;); \neg; \vee\}$$

by

1. Any one of the variables $p, p', q, \ldots$ and constants *false*, *true* is a formula.
2. If A is a formula, then so is $(\neg A)$
3. If A and B are formulas, then so is $(A \vee B)$

We say that B is a *subformula* of a formula A iff B is a *substring* of A, and, moreover it is a formula.

For example, p, q, $(\neg p)$, $((\neg p) \vee q)$ are all subformulas of $((\neg p) \vee q)$.

None of $\neg$, $\neg p$, $p) \vee q$ are subformulas, since none is a formula—even though they are substrings of $((\neg p) \vee q)$.

A *proper* subformula of A is a subformula that is not the exact same string as A. The first three examples of the subformulas above are proper, but the fourth is not.

With the definitions out of the way we observe that $\prec$ in this context is well-founded, so it has IC. This allows us to prove properties of formulas (that is, members of X) by GCVI with respect to $\prec$.

Wait! Why is $\prec$ well-founded? Because starting with A we cannot have an infinite chain $\ldots A'' \prec A' \prec A$ since every application of $\prec$ as we walk from right to left removes at least one symbol from A. But A has finitely many symbols!

Let us now be of some use. We will prove:

Every Boolean formula has an equal number of left and right brackets.

Proof. We are asked to prove the "property", as stated above, for every formula $A \in X$. We use GCVI (see 1.1 and the discussion following).

(1) (I.H.) Assume the claim for all $B \prec A$.

(2) (Goto) Prove the claim for A. We have cases:

Case 1. A is $p, q, \ldots$ or *false*, or *true*. This is the basis, or "boundary" case these atomic formulas have no proper subformulas and thus the proof is not helped by the I.H.
In each of those cases the claim is true (0 left and 0 right brackets).

Case 2. A is $(\neg B)$. By I.H., since $B \prec A$, B has as many left as it has right brackets. This is true of A as well, since it has one extra left and one extra right bracket than B has.

Case 3. A is $(B \vee C)$. Since $B \prec A$ and $C \prec A$, the I.H. applies to both B and C. Let B have m left and m right brackets and let n be the corresponding number for C. Then A has $1 + m + n$ left, and as many right brackets. ■

The sentence, in the basis case above, "these atomic formulas have no proper subformulas" can be rephrased: "These atomic formulas are the minimal elements of X" since, for example, $p \in X$ but for *no* $z \prec p$ is it the case that also $z \in X$.

Contrast this with $\mathbb{N}$ and the standard order relation $<$. Minimal elements of any S—where $\emptyset \neq S \subseteq \mathbb{N}$—are unique.

But there is another difference between the pairs $\prec, X$ and $<, \mathbb{N}$:

(1) $<$ satisfies trichotomy (3rd Peano axiom for $<$).

(2) $\prec$ does not satisfy trichotomy *in general*—meaning that there are *specific* concrete instances of the abstract $\prec$ that do not satisfy trichotomy. The specific concrete version in the previous example is one such. For example, compare two distinct Boolean variables p and q. First off, they are different strings, so $p = q$ (equality of strings, informally!) is false.

Pause. Why "informally"?

But so are $p \prec q$ and $q \prec p$.

Thus, we *cannot* say that

$$(\forall x)(\forall y)(x \prec y \vee x = y \vee y \prec x) \tag{3}$$

is true <u>for this $\prec$</u> (proper subformula relation). Therefore we cannot say this *in general* either.

There is a connection between uniqueness of minimal elements and trichotomy for our well-founded $\prec$:

First, whenever trichotomy holds for some concrete choice of $\prec$—as it does when we take $\prec$ to be $<$ on $\mathbb{N}$—then a minimal element is also *minimum* or *least*.

Indeed, if $a \in S \subseteq X$ is minimal, then we have $\neg(\exists z \prec a) z \in S$ by definition. This is short for $\neg(\exists z)(z \prec a \wedge z \in S)$, which is logically equivalent to $(\forall z)(\neg z \in S \vee \neg z \prec a)$. Now, having assumed trichotomy (3) allows us to conclude

$$\neg z \prec a \equiv z = a \vee a \prec z \tag{4}$$

Indeed, the $\Rightarrow$ is a tautological consequence of a specialisation of (3). For $\Leftarrow$ we use well-foundedness: Assume $z = a \vee a \prec z$. Now argue by contradiction, and assume also $z \prec a$. This along with our assumption gives (written conjunctionally) $z \prec a = z$ or $z \prec a \prec z$. Each of these two cases contradicts well-foundedness.

Thus the element a satisfies $(\forall z)(z \in S \Rightarrow a \prec z \vee a = z)$. In words, every member of S is above a or is equal to a. For short, a is $\prec$-*least* in S. Now least members are unique:

Indeed, suppose that a and b are both least (with respect to the relation $\prec$) in the nonempty subset S of X. By way of contradiction let $a \neq b$. Then $a \prec b$ (taking a as least) and $b \prec a$ (taking b as least). This yields $a \prec b \prec a$ contradicting the well-foundedness of $\prec$.

By the way, you see now why MC for $<$ on $\mathbb{N}$ is called the *least principle*: *Minimal elements in this context are also least.*

Secondly, uniqueness of minimal elements implies trichotomy. This is easy: Suppose $\prec$ on X has MC and that moreover every $\emptyset \neq S \subseteq X$ has *no more than one*[7] minimal element.

I want to show that any two x and y in X are "comparable", that is, one of $x = y$, $x \prec y$ or $y \prec x$ holds. Well, if none of $x = y$, $x \prec y$ and $y \prec x$ is true, let us form the set $S = \{x, y\}$. This set has two distinct minimal elements, contrary to our assumption.

Example 2.2. We continue with the X and $\prec$ of example 2.1. This time we prove:

[7] MC guarantees at least one.

Every nonempty proper prefix of a formula A has more left brackets than right brackets.

By "prefix" we mean a substring that starts with the first symbol of A. It is *proper* if it is not the same string as A.

Proof. By GCVI.

(1) (I.H.) Assume the claim for all $B \prec A$.

(2) (Goto) Prove the claim for A. We have cases:

Case 1. A is $p, q, \ldots$ or *false*, or *true*. None of the minimal elements of X has any nonempty proper prefix. So there is nothing to prove.

Case 2. A is $(\neg B)$. Here are all the possible nonempty proper prefixes (enclosed in quotes):

- (a) "(" This has the property.
- (b) "$(\neg$" Ditto.
- (c) "$(\neg R$", where R is a proper nonempty prefix of B. Since $B \prec A$, the I.H. tells us that R has more lefts than rights. Adding the leftmost "(" to that does no harm.
- (d) "$(\neg B$" B has equal number of lefts and rights (by example 2.1). But "$(\neg B$" has an extra "(" as its first symbol.

Case 3. A is $(B \vee C)$. Here are all the possible nonempty proper prefixes (enclosed in quotes):

- (a) "(" This has the property.
- (b) "$(R$", where R is a nonempty proper prefix of B. Since $B \prec A$, the I.H. tells us that R has more lefts than rights. Adding the leftmost "(" to that does not harm the balance.
- (c) "$(B$" B has equal number of lefts and rights (by example 2.1). But "$(B$" has an extra "(" as its first symbol.
- (d) "$(B\vee$" Exactly as in the previous case.
- (e) "$(B \vee T$", where T is a nonempty proper prefix of C. Since $C \prec A$, the I.H. tells us that T has more lefts than rights. Adding the leftmost "(" to that, and remembering that B has as many lefts as rights, does not harm the balance.
- (f) "$(B \vee C$" B and C each have as many lefts as rights by 2.1. The leftmost "(" saves the day. ■

3 Binary Trees

The next example of application of GCVI is significant enough to deserve its own section. We will look into the definition of what we call *extended binary trees* and prove some of their properties by GCVI with respect to the appropriate $\prec$ that we will define shortly. But first let us define these trees. We have an infinite supply of *labelled* squares and *labelled* circles—$\Box$ and $\bigcirc$—that we will call square and round *nodes* respectively. We also have an infinite supply of line segments that we call *edges*. We usually do not display the labels of nodes unless we want to talk about a particular node by referring to it via its label.

Definition 3.1 (Extended Binary Trees). An *Extended Binary Tree*, its *root* and its *support* are defined simultaneously by induction. We will simply say "tree" in what follows.

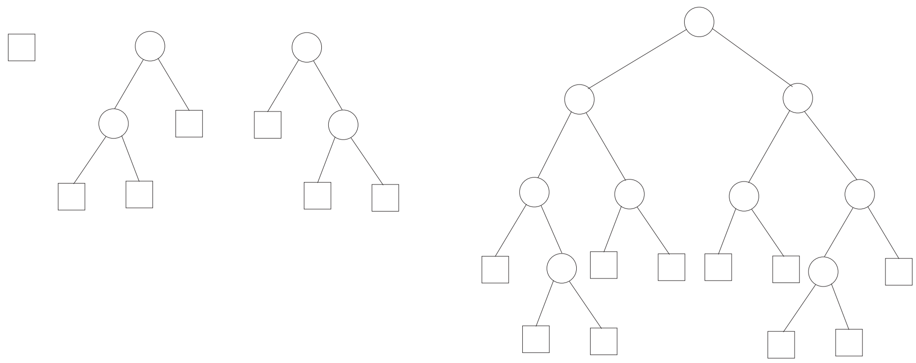
The support will be the set of nodes used to build the tree. We write $sup(T)$ to indicate the support of tree T. The root is a special node of the tree. We write $root(T)$ to indicate the root of T.

A tree is one of:

(1) A single square node, $\Box$. In this case $root(\Box) = \Box$ and $sup(\Box) = \{\Box\}$.

(2) A structure formed as follows: We start with two trees T_1 and T_2 such that $sup(T_1) \cap sup(T_2) = \emptyset$. We get a round node, say r (its label), that is *not* a member of $sup(T_1) \cup sup(T_2)$ and two unused edges. We connect $root(T_1)$ to the *left* of r using one edge, and connect $root(T_2)$ to the *right* of r using the other edge.
Order matters!
The ordered triple so formed, (T_1, r, T_2) is a tree. Let us call it T.
By definition, $root(T) = r$ and $sup(T) = sup(T_1) \cup sup(T_2) \cup \{r\}$.
We say that T_1 is the *left subtree* of T (or of r) and T_2 is the *right subtree* of T (or of r). ■

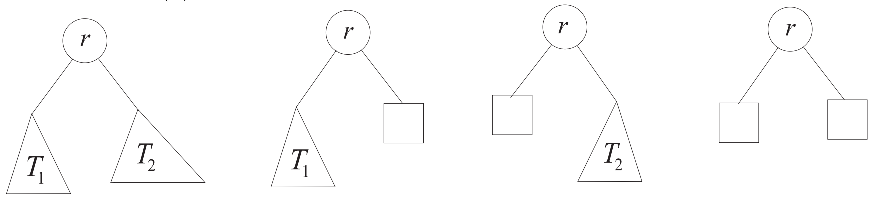
Example 3.2. We present four examples of trees



Note that the 2nd and 3rd examples above are different because order matters. In terms of computer science practice, you may think of trees as data structures that hold information. The data is stored in round notes. The edges are "links" (pointers) that point from top to bottom. The square nodes indicate "null addresses" thus the edges pointing to them are "null links". Therefore, square nodes never hold data. ■

In general we draw an arbitrary tree as a big "Δ" when we do not care to display the tree in detail. In such drawings we always imply that the root is at the top tip of the letter. We may use this "compressed" drawing for an entire tree or for parts of it.

For example, as in the pictures of the four trees below, where some subtrees are denoted by big Δ's. The first tree displayed is the "general tree" of case (2) in definition 3.1.



Note that in much of the literature square nodes are called *external* or *leaves* (leaf); round nodes are called *internal*.

We have used the terms *left subtree* and *right subtree* of a tree T. Take the left subtree, T_1, of T. If it is not just a $\Box$ it has a left and a right subtree. And so on.

What will we call the left subtree of the right subtree of the left subtree of T? Let us just call it "a subtree of T".

In general,

Definition 3.3 (Subtrees). We define a relation $\prec$ on the set X of all trees as follows:

We write $S \prec T$—pronounced "S is a subtree of T"—to mean that there is a finite sequence $T_1, T_2, \ldots, T_n$ (where for convenience we have renamed S into T_1 and T into T_n) with the property that for all i such that $1 \le i < n$, T_i is the left or right subtree of T_{i+1}. ■

Whenever $S \prec T$, we have $sup(S) \subset sup(T)$ since at the very least $root(T) \notin sup(S)$. Thus every chain of trees

$$\ldots T''' \prec T'' \prec T' \prec T$$

terminates because walking towards the left, sooner or later we run out of nodes.

That is $\prec$ on X is well-founded and therefore has MC and IC. In particular, we can prove properties of trees by GCVI.

Example 3.4. Prove that a tree T of $n+1$ nodes has n edges.

Pause. Why $n+1$? Because by definition, every tree, even the smallest one (case (1) in definition 3.1) has at least one node.

Proof.

(1) (I.H.) Assume that this is true for all $S \prec T$.

(2) (Goto) Prove for T. We have cases:

1. $T = \square$ ("boundary case") Clearly we have one node and 0 edges by 3.1.
2. T is the ordered triple (T_1, r, T_2) (refer if you want to the first, "the general", tree displayed in the previous figure—but this is not necessary).
   Since $T_1 \prec T$ and $T_2 \prec T$, the I.H. applies to both of T_1 and T_2.
   Say the first has e_1 edges and e_1+1 nodes, while the second has e_2 edges and e_2+1 nodes.
   Now T has e_1+e_2+2 edges and $1+(e_1+1+e_2+1)$ nodes, where the first "$1+$" accounts for the root r. A comparison of these numbers shows that we are done. ■

The following corollary is a useful tool in the analysis of searching and sorting algorithms.

Corollary 3.5. *The number of square nodes in a tree is exactly one more than the round nodes.*

Proof. Let T be a tree, and let it have $\mathscr{R}$ round and $\mathscr{S}$ square nodes. By the above example it has $\mathscr{R} + \mathscr{S} - 1$ edges.

However, the number of edges is also $2\mathscr{R}$ (you see why?). Thus,

$$\mathscr{R} + \mathscr{S} - 1 = 2\mathscr{R}$$

The above is equivalent to $\mathscr{S} - 1 = \mathscr{R}$. ■

There are "tall" trees, and then there are "short" trees:

Definition 3.6 (Height of a tree). One defines the height of a tree T by induction on the definition of trees (3.1).[8] We write $h(T)$ for the height of T.

If $T = \square$, then $h(T) = 0$.

If T is the ordered triple (T_1, r, T_2), then $h(T) = 1 + \max\big(h(T_1), h(T_2)\big)$. ■

Example 3.7. The heights of the trees displayed in example 3.2 are, from left to right, $0, 2, 2, 4$. ■

You surely did not miss to observe that the height of a tree equals the number of edges in a longest path from root to a square node.

Thus, if the tree implements a so-called "sort-tree" (that you must have seen in COSC 2011), then its height represents the worst case number of "probes" (i.e., comparisons with round-note data) when you search for an item that is *not* stored in the tree (unsuccessful search).

Example 3.8. Prove that if a tree T has $n + 1$ nodes and height h, then

$$n + 1 \leq 2^{h+1} - 1 \tag{1}$$

We use GCVI.

(I) (I.H.) Assume that this is true for all $S \prec T$.

[8] It is "very fine print" to explain why such inductive definitions do manage to define what we think they define. So we will sweep the reasons under the rug and take the existence and uniqueness of such inductively defined objects—here the tree heights—for granted.

(II) (Goto) Prove for T. We have cases:

1. $T = \Box$. Here $h(T) = 0$, and $n + 1 = 1$. Thus (1) holds ($2^1 - 1 = 1$).
2. T is the ordered triple (T_1, r, T_2).
Since $T_i \prec T$ for $i = 1, 2$, I.H. applies. Let T_i have $n_i + 1$ nodes and height h_i. By I.H.
$$n_1 + 1 \leq 2^{h_1+1} - 1 \tag{2}$$
and
$$n_2 + 1 \leq 2^{h_2+1} - 1 \tag{3}$$
T has $n + 1 = n_1 + n_2 + 3$ nodes. Adding (2) and (3) we have
$$n_1 + n_2 + 2 \leq 2^{h_1+1} + 2^{h_2+1} - 2 \leq 2 \cdot 2^{h_1+1} - 2 = 2^{h_1+2} - 2 \tag{4}$$
where, without loss of generality, we assumed that $h_1 \geq h_2$, that is, $h_1 = \max(h_1, h_2)$.
(4) can be rewritten as
$$n_1 + n_2 + 3 \leq 2^{h_1+2} - 1$$
and we are done, for lhs $= n + 1$ and rhs $= 2^{h(T)} - 1$. ■

This result says that the worst number of probes for unsuccessful search in a sort-tree cannot be less than $\log_2(n + 2) - 1$.