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

1 Special Relations; Relational closures

We continue within informal mathematics until otherwise stated.[1]

We will continue for a while looking only at relations $S : A \to A$, however the definition below applies to any relations, possibly with distinct left and right fields. Indeed, *the definition is independent of the fields.*

Definition 1.1 (Relational Inverse). For any relation R, we define

$$R^{-1} \stackrel{\text{Def.}}{=} \{\langle x, y\rangle | yRx\} \tag{1}$$

We call R^{-1} the inverse of R. ■

Of course, the definition could have been given as

$$(\forall x)(\forall y)(xR^{-1}y \equiv yRx)$$

a fact that is equivalent to (1). As it is usual, one omits the quantifiers (in one direction by specialisation, in the other by—the allowed in set theory—generalisation) and writes:

$$xR^{-1}y \equiv yRx$$

Clearly, $(R^{-1})^{-1} = R$. Indeed,

$$\begin{aligned}
&x(R^{-1})^{-1}y \\
\equiv &\left\langle 1.1 \right\rangle \\
&yR^{-1}x \\
\equiv &\left\langle 1.1 \right\rangle \\
&xRy
\end{aligned}$$

[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.

Exercise 1.2. Prove that $(R \cup S)^{-1} = R^{-1} \cup S^{-1}$. ■

Exercise 1.3. Prove that $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$. ■

Remark 1.4. We defined $\mathrm{dom}(R) = \{x|(\exists y)xRy\}$ and $\mathrm{ran}(R) = \{y|(\exists x)xRy\}$ in Part IV. Since we have $(\exists y)xRy \equiv (\exists y)yR^{-1}x$ by sWLUS, we get

$$\mathrm{dom}(R) = \{x|(\exists y)xRy\} = \{x|(\exists y)yR^{-1}x\} = \mathrm{ran}(R^{-1})$$

Similarly,

$$\mathrm{dom}(R^{-1}) = \mathrm{ran}(R)$$

In particular, R is total iff R^{-1} is onto and R is onto iff R^{-1} is total. ■

There is a number of relation types that are of interest:

Definition 1.5. Given a relation $R : A \to A$.

1. It is *reflexive* iff $(\forall x \in A)xRx$.
2. It is *irreflexive* iff $(\forall x)\neg xRx$
3. It is *symmetric* iff $(\forall x)(\forall y)(xRy \Rightarrow yRx)$
4. It is *antisymmetric* iff $(\forall x)(\forall y)(xRy \wedge yRx \Rightarrow x = y)$
5. It is *transitive* iff $(\forall x)(\forall y)(\forall z)(xRy \wedge yRz \Rightarrow xRz)$ ■

Only part 1 of the definition needs to refer to A. Indeed it depends very much on it. Consider $R = \{\langle 1, 1\rangle\}$. If $A = \{1\}$, then R is reflexive. If $A = \{1, 2\}$, then it is not reflexive, because now it should have the pair $\langle 2, 2\rangle$ in it, but it does not.

R is all of 3–5 regardless of A. An example of an irreflexive relation is $\{\langle 1, 2\rangle\}$. Other examples are $<$ on $\mathbb{N}$ and $\subset$ on sets. Examples of antisymmetric relations, beyond the particular R of this example, are $\leq$ on $\mathbb{N}$ and $\subseteq$ on sets (by extensionality).

Note that

$$(\forall x)(\forall y)(xRy \Rightarrow yRx) \equiv (\forall y)(\forall x)(yRx \Rightarrow xRy) \equiv (\forall x)(\forall y)(yRx \Rightarrow xRy)$$

where the first $\equiv$ is by dummy renaming (and WLUS) and the second by commuting the $\forall$'s. Thus we get the $\Leftarrow$ direction in 3 for free, and we have the theorem "R is symmetric iff $(\forall x)(\forall y)(xRy \equiv yRx)$". Actually, we have just proved the "only if" ($\Rightarrow$) direction. The "if" direction is by $\forall$-MON and the tautology $(A \equiv B) \Rightarrow (A \Rightarrow B)$.

Example 1.6. $R : A \to A$ is given. Then

(i) R is reflexive iff $1_A \subseteq R$.
(ii) R is symmetric iff $R = R^{-1}$.
(iii) R is irreflexive iff $R \cap 1_A = \emptyset$.
(iv) R is transitive iff $R^2 \subseteq R$.
(v) R is antisymmetric iff $R \cap R^{-1} \subseteq 1_A$.

Let us do (i) and (ii) and leave the rest as exercises.

(i): <u>Assume $1_A \subseteq R$.</u>

Now prove that R is reflexive. So let $x \in A$ and prove xRx. Since $1_A = \{\langle x, x\rangle | x \in A\}$ by definition (Part IV, 3.5), we have $x1_Ax$. By the hypothesis, I have xRx. Done.

For the other direction, <u>assume that R is reflexive.</u>

Prove $1_A \subseteq R$. Well, <u>let</u> $x1_Ax$. By definition of identity, $x \in A$. By definition of reflexivity and by the hypothesis, xRx. Connecting with our "let", we have what we want.

(ii): <u>Assume that $R = R^{-1}$.</u>

I want[2] $xRy \equiv yRx$ (remember: I can place the universal quantifier afterwards). Well, $xRy \equiv yR^{-1}x \equiv yRx$, where the 2nd $\equiv$ is by hypothesis.

Conversely, <u>assume that $(\forall x)(\forall y)(xRy \equiv yRx)$</u>. Thus, $(\forall x)(\forall y)(xRy \equiv xR^{-1}y)$ by sWLUS and definition of inverse. Therefore (by extensionality) $R = R^{-1}$. ■

We now turn to "closing" relations. I get a *closure of R with respect to a property* (such as reflexivity, symmetry, etc.) by adding *just enough, <u>but no more than needed</u>* pairs to R so as to make it have the required property. Rigourously then, we define

Definition 1.7 ("Popular" closures). We are back to relations on a set A. So let $R : A \to A$ be given.

(a) The *reflexive closure* of R is the $\subseteq$-smallest reflexive S such that *extends* R, i.e., $R \subseteq S$.
We write $S = r(R)$.
(b) The *symmetric closure* of R is the $\subseteq$-smallest symmetric S such that *extends* R, i.e., $R \subseteq S$.
We write $S = s(R)$.

[2] You haven't forgotten what I proved just before this example, have you?

(c) The *transitive closure* of R is the $\subseteq$-smallest transitive S such that *extends* R, i.e., $R \subseteq S$.
We write $S = t(R)$ or $S = R^+$. ■

Let us "parse" the definition. Of course, the "extends" part is $R \subseteq S$. What do I mean by $\subseteq$-smallest? Intuitively, I mean that I do not add to R more pairs than I need to make R one of reflexive, symmetric, transitive.

Formally the definition translates into:

(1) The *reflexive closure* of R is a relation S such that
   (a) $R \subseteq S$ (the extends part)
   (b) S is reflexive
   (c) If $R \subseteq T$ and T is also reflexive, then $S \subseteq T$. This is the "$\subseteq$-smallest" part. That is, any other reflexive extension is equal to or larger than the closure ("larger" meaning "superset").

Similarly,

(2) The *symmetric closure* of R is a relation S such that
   (a) $R \subseteq S$
   (b) S is symmetric
   (c) If $R \subseteq T$ and T is also symmetric, then $S \subseteq T$.
(3) The *transitive closure* of R is a relation S such that
   (a) $R \subseteq S$
   (b) S is transitive
   (c) If $R \subseteq T$ and T is also transitive, then $S \subseteq T$.

Remark 1.8 (Uniqueness of closures). We have used the definite article in the definition of closures, 1.7. This is justified, because any of the three closures we defined for a relation R is *unique if it exists*.

We will worry about existence shortly. Uniqueness is easy.

Let S be a reflexive (symmetric, transitive) closure of R, and let also T be another one.

So, each of S and T extend R, and each is reflexive (symmetric, transitive).

Since S is $\subseteq$-smallest such, we have $S \subseteq T$. But T is also smallest such, because we assumed it is a closure too. That is, $T \subseteq S$. By extensionality, $S = T$. ■

Example 1.9. Thus, if $R = \{\langle 1,2\rangle, \langle 2,3\rangle, \langle 3,1\rangle\}$, then

$$r(R) = \{\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle, \langle 1,2\rangle, \langle 2,3\rangle, \langle 3,1\rangle\}$$

$$s(R) = \{\langle 2,1\rangle, \langle 3,2\rangle, \langle 1,3\rangle, \langle 1,2\rangle, \langle 2,3\rangle, \langle 3,1\rangle\}$$

and

$$t(R) = \{\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle, \langle 1,2\rangle, \langle 2,3\rangle, \langle 3,1\rangle, \langle 2,1\rangle, \langle 3,2\rangle, \langle 1,3\rangle\}$$

Lest you think that $t(R)$ always ends up symmetric and reflexive, here is a counterexample: Start with $S = \{\langle 1,2\rangle, \langle 2,3\}$. Then

$$t(S) = \{\langle 1,2\rangle, \langle 2,3\rangle, \langle 1,3\rangle\}$$

■

We settle the existence of closures *constructively*, by showing how to compute them:

Theorem 1.10 (Existence of closures). *For any relation $R : A \to A$,*

(1) $r(R) = 1_A \cup R$
(2) $s(R) = R \cup R^{-1}$
(3) $t(R) = \bigcup_{i \ge 1} R^i$

Before we embark with the proof, let me explain the symbol

$$\bigcup_{i\ge 1} R^i \tag{4}$$

It means, intuitively,

$$R \cup R^2 \cup R^3 \cup \ldots \cup R^i \cup \ldots \text{ without end}$$

That is, $\langle x, y\rangle$ is in (4) iff it is in at least one of the *positive* powers R^i. Formally then, it means

$$\bigcup_{i\ge 1} R^i = \{\langle x, y\rangle | (\exists i \ge 1) x R^i y\} \tag{5}$$

We can now turn to the proof of the theorem.

Proof. (1) Trivially, $R \subseteq 1_A \cup R$. Also, $1_A \cup R$ is reflexive by example 1.6. We need to show that our "solution" is smallest. Let then also

$$R \subseteq T \tag{6}$$

and T be reflexive. Again by 1.6, $1_A \subseteq T$ which by (6) gives $1_A \cup R \subseteq T$.

(2) Trivially $R \subseteq R \cup R^{-1}$. Moreover, our proposed solution is symmetric by 1.6 and exercise 1.2. Here is the contribution of exercise 1.2:

$$(R \cup R^{-1})^{-1} = R^{-1} \cup (R^{-1})^{-1} = R^{-1} \cup R$$

The last "=" is by the remark following 1.1.

To see that the proposed solution works I must show it is smallest. Let then $R \subseteq T$ and T be symmetric. Clearly $R^{-1} \subseteq T^{-1}$,[3] hence $R^{-1} \subseteq T$ since $T = T^{-1}$ by example 1.6. All told, $R \cup R^{-1} \subseteq T$.

(3) Let us call $\bigcup\limits_{i \geq 1} R^i$ "S". Clearly, $R \subseteq S$, since $S = R \cup R^2 \cup \ldots$[4]

Next we show that S is transitive, so let $xSySz$. Thus I have

$$(\exists i \geq 1) x R^i y \tag{7}$$

and

$$(\exists i \geq 1) y R^i z \tag{8}$$

Informally, let $i = k$ work for (7) and $i = m$ work for (8), so we have[5]

$$k \geq 1 \wedge x R^k y \tag{7$'$}$$

and

$$m \geq 1 \wedge y R^m z \tag{8$'$}$$

(7′) and (8′) yield $k + m \geq 1 \wedge x R^k \circ R^m z$, hence, using proposition 3.10 of Part IV,

$$k + m \geq 1 \wedge x R^{k+m} z$$

Just as in footnote 4, the above yields $(\exists i \geq 1) x R^i z$, i.e., xSz.

---

[3]Conjunctionally, $xR^{-1}y \Rightarrow yRx \overset{\text{hypothesis}}{\Rightarrow} yTx \Rightarrow xT^{-1}y$.

[4] A formal Hilbert proof without the numbering goes like this: Let xRy. Then xR^1y since $R = R^1$. Since $1 \geq 1$ is a theorem—that is, $1 = 1 \vee 1 < 1$—I now have $1 \geq 1 \wedge xR^1y$. Apply now the rule $A[x := t] \vdash (\exists x)A$ to get $(\exists i)(i \geq 1 \wedge xR^iy)$, or for short $(\exists i \geq 1)xR^iy$. But that says $x \bigcup\limits_{i \geq 1} R^i y$, i.e., xSy.

[5]Formally, we say the same thing like this: Let k be a fresh variable and assume (7′). Moreover, let m be a fresh variable and assume (8′). Note how the "freshness" for you against the error of choosing the same "value" of i both (7′) and (8′). The informal proof speaks of "values", the formal speaks of "names". One tracks the other faithfully.

Finally! Let us prove that our "solution" S is the smallest transitive extension of R. So let T be another transitive extension:

$$R \subseteq T \text{ and } T \text{ is transitive} \tag{9}$$

It suffices to prove that

$$(\forall i \geq 1)(R^i \subseteq T) \tag{10}$$

for then T is a superset of each set in the union

$$R \cup R^2 \cup R^3 \cup \ldots$$

and therefore a superset of the union itself.

So let us prove (10) by induction on i. The basis $i = 1$ is the assumption (9).

So assume $R^i \subseteq T$ (I.H.)

We next goto $i+1$: Let $xR^{i+1}y$. By definition of powers, this means $xR \circ R^i y$. Hence for some z (formally this would be a fresh variable) I have xRz and $zR^i y$. The first of these two conclusions gives xTz by (9). The second gives zTy by I.H. Since T is transitive, I got xTy and I am done. ∎

In class we formalised the part "for then T is a superset of each set in the union

$$R \cup R^2 \cup R^3 \cup \ldots$$

and therefore a superset of the union itself". Can you reproduce that formal proof, which was based on the formal definition of $\bigcup_{i \geq 1} R^i$?

What does xR^2y say intuitively? That R allows us to go from x to y in two R-steps, since there must be a z such that $xRzRy$. Similarly, xR^3y says that we can go from x and y in 3 steps, and, in general, xR^ny says that we can go from x to y in n steps. No wonder then that

$$R^+ = R \cup R^2 \cup R^3 \cup \ldots$$

for the first term, R, is what we start with. The 2nd, R^2 adds to R those pairs that allow one to bridge x, y in one step, whereas without this addition it may have (if R is not transitive) taken two steps. Similarly, the pairs that R^3 adds are those x, y that originally we could bridge in 3 steps. Adding the pair outright means that I can also go from x, y in one step, as transitivity would require. Think about it: If R were transitive to begin with, would it not be that $xRzRwRy$ implies xRy? That is, along with the 3-step route there *must* be a "direct" route as well? The infinite union that "computes" R^+ ensures that all direct routes are added.

One more thought: xR^+y is true iff I can get from x to y in one or more R-steps. Indeed, xR^+y says just that: $(\exists i \geq 1)xR^iy$.

Remark 1.11. One often uses $r(t(R))$. This turns out to be equal to $t(r(R))$. The literature usually uses the symbol R^* for either, and calls it the *reflexive transitive closure* of R. It can be computed by

$$R^* = 1_A \cup R \cup R^2 \cup \ldots = \bigcup_{i \geq 0} R^i$$

Clearly, assuming that the above computation for R^* is correct, xR^*y is true iff I can get from x to y in zero or more R-steps. Zero steps means $x = y$ (from xR^0y or $x1_Ay$). ■

2 Equivalence Relations

A relation $R : A \to A$ that is all of reflexive, symmetric and transitive is called an *equivalence relation*. These play a major role in computer science and mathematics. For example, a practical application is in the minimisation of finite automata (a topic that may be found in COSC2001 or in COSC3302).

Example 2.1. Any 1_A and the "$\equiv$" of logic are equivalence relations.

Here is a more interesting one on $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. For any $m > 1$ define the relation $\equiv_m$ defined by:

$$a \equiv_m b \quad \text{iff} \quad m \text{ is a factor of } a - b \tag{1}$$

A number theorist will probably rather write "$a \equiv_m b$" as "$a \equiv b \pmod{m}$" or "$a \equiv b\ (m)$".

In any case we pronounce "$a \equiv_m b$"—by this or any other notation—"*a is congruent to b modulo m*".

Enough jargon. Let us verify that $\equiv_m$ is indeed an equivalence relation on $\mathbb{Z}$.

Reflexivity is trivial, for $a - a = 0$ and m is certainly a factor of 0. For symmetry, say $a \equiv_m b$. Thus m is a factor of $a - b$. But then so is of $b - a$.

Finally, let $a \equiv_m b \equiv_m c$. Let us rewrite the hypothesis in "factor of" notation. So

$$a - b = m \cdot k$$

and

$$b - c = m \cdot n$$

for some k, n. Add and get $a - c = m \cdot (k + n)$. Done.

By the way, this is an equivalence relation that is not antisymmetric. Indeed, for any fixed m, $0 \equiv_m m$ and $m \equiv_m 0$, yet $0 \neq m$ (recall that $m > 1$). Note that 1_A *is* antisymmetric. However the $\equiv$ of logic is not. For example, $p \equiv (p \wedge p)$ and $(p \wedge p) \equiv p$, but $p \neq (p \wedge p)$ (as strings, that is, they are different; equivalence does not force them to be the same). ■

The following concept is important:

Definition 2.2 (Equivalence classes). Let $R : A \to A$ be an equivalence relation. We define for each $x \in A$ a special set that we call "*the equivalence class of R represented by x*". The symbol is $[x]_R$, where we may omit the subscript if the context makes it clear.

$$[x]_R \stackrel{\text{Def.}}{=} \{y \in A | yRx\}$$

■

It is immediate that $a \in [a]_R$ since reflexivity gives aRa.

If all the relations in a given discussion are on the same set A, then we may omit the obvious "$y \in A$" above and write instead $[x]_R = \{y | yRx\}$

Example 2.3. The equivalence classes "modulo 2" are the sets $[x]_{\equiv_2}$. It is easy to verify that there are only two classes, one with representative 0 and one with representative 1. The first contains all the even numbers the second contains all the odd numbers. ■

Hey wait a minute! I think that I can represent all the even numbers using 2 as the representative, i.e., $[0]_{\equiv_2} = [2]_{\equiv_2}$.

This is not an accident:

Lemma 2.4. *Let $R : A \to A$ be an equivalence relation. Then (for all a, b) aRb iff $[a] = [b]$.*

Proof. ($\Rightarrow$) <u>Assume aRb</u>.

We want $[a] = [b]$. Towards that,

($\subseteq$) Let $x \in [a]$. Hence (def. 2.2), xRa. The underlined assumption and transitivity yields xRb, hence $x \in [b]$.

($\supseteq$) Let $x \in [b]$. Hence xRb. Since the underlined assumption and symmetry gives bRa, transitivity now yields xRa, hence $x \in [a]$.

($\Leftarrow$) $\underline{\text{Assume } [a] = [b]}$.

We want aRb. Well, $a \in [a]$, hence, by assumption, $a \in [b]$. Definition 2.2 yields aRb. ■

Thus any $x \in [a]$ is as good as a in the job of representative. Indeed, the assumption yields xRa by 2.2, hence $[x] = [a]$ by the lemma.

We can now easily prove:

Theorem 2.5. *Let $R : A \to A$ be an equivalence relation. Then*

(1) *If $x \in A$, then* $[x] \neq \emptyset$
(2) *If x and y are in A, then* $[x] \cap [y] \neq \emptyset \Rightarrow [x] = [y]$
(3) $\bigcup_{x \in A} [x] = A$.

Proof. (1) By $x \in [x]$ (see remark following definition 2.2).

(2) Assume $[x] \cap [y] \neq \emptyset$. So let $z \in [x] \cap [y]$. Thus zRx and zRy. The 1st of these conclusions yields xRz by symmetry. Along with the second and transitivity I get xRy. By lemma 2.4 I now have $[x] = [y]$ as needed.

(3) I get $\bigcup_{x \in A} [x] \subseteq A$ trivially, since for any $x \in A$ I have $[x] \subseteq A$ by definition 2.2. For $\supseteq$ note that $[x] \supseteq \{x\}$. Taking unions on both sides I have $\bigcup_{x \in A} [x] \supseteq \bigcup_{x \in A} \{x\}$. But $\bigcup_{x \in A} \{x\} = A$. ■

Abstracting properties (1)–(3) of equivalence classes one defines partitions of sets:

Definition 2.6 (Partitions). A *family of subsets* of a set A is a *partition of (or "on") A* iff F satisfies:

(1) $(\forall S \in F) S \neq \emptyset$
(2) $(\forall S \in F)(\forall T \in F)(S \cap T \neq \emptyset \Rightarrow S = T)$
(3) $\bigcup F = A$.

Sometimes the members of the partition, i.e., the various S in F, are called *blocks*. ■

Example 2.7. So, if $R : A \to A$ is an equivalence relation, then $F = \{[x] | x \in A\}$ is a partition on A.

By the way, this kind of partition that arises from an equivalence relation is often denoted by A/R. ■

Actually, partitions are not more general than sets of equivalence classes as the following shows.

Theorem 2.8. *Let F be a partition on A. Define a relation $R : A \to A$ by*

$$aRb \overset{Def.}{\equiv} (\exists S \in F)(a \in S \land b \in S)$$

Then

(1) *R is an equivalence relation*
(2) *$A/R = F$*

Proof. (2) I leave as an interesting exercise (Problem set #5). Let me do (1):

Reflexivity: So let $a \in A$. By property (3) in definition 2.6, there is an $S \in F$ such that $a \in S$. In symbols, $(\exists S \in F)(a \in S \land a \in S)$ holds. This says aRa.

Symmetry: So let aRb. Thus, $(\exists S \in F)(a \in S \land b \in S)$ by the definition of R. sWLUS now gives $(\exists S \in F)(b \in S \land a \in S)$, that is, bRa.

Transitivity: So let aRb and bRc. The definition of R gives

$$(\exists S \in F)(a \in S \land b \in S) \tag{$*$}$$

and

$$(\exists S \in F)(b \in S \land c \in S) \tag{$**$}$$

Let $S = T$ work in $(*)$, that is,

$$a \in T \land b \in T \tag{I}$$

Let $S = W$ work in $(**)$, that is,[6]

$$b \in W \land c \in W \tag{II}$$

Since $b \in T \cap W$, property (2) of F (see 2.6) yields $T = W$. Thus, $a \in T \land c \in T$ from $(I), (II)$. So an $S \in F$ exists—take $S = T$—such that $a \in S \land c \in S$. For short, aRc, as needed. ∎

[6] We have no *a priori* right to say "let $S = T$", i.e., to use the same "value" for S in both $(*)$ and $(**)$. Of course, if we later prove them equal, that is fine. Note how the formal approach protects us from "letting" $S = T$ in both cases, because formally we eliminate $\exists$ from $(*)$ and introduce a fresh variable T in the place of S. When we eliminate the second $\exists$ we are again obliged to get a new variable, one that has not been used yet.

3 Functions

Functions, *intentionally* are agents ("devices") that receive inputs, and for each input return *at most one* output. Extensionally then they are nothing but relations, i.e., sets of in/out *pairs*,[7] except for the important restriction of "single-valued-ness" of output, the "at most" qualification.

We define below which relations are functions. We return our attention to relations $R : A \to B$ where $A \neq B$, in general, for the balance of Part V. Moreover, for the balance of this section "function" is exclusively an extensional object, a set of ordered pairs, as defined below.

Definition 3.1 (Functions). A relation $f : A \to B$ is a function iff it is *single-valued in the 2nd projection*, that is,

$$(\forall x)(\forall y)(\forall z)(xfy \wedge xfz \Rightarrow y = z)$$

You noticed the "f". Generically, we will denote functions by f, g, h, with primes and/or subscripts if we run out of letters. ■

Informally, one could have just implied the quantifiers and written instead "$xfy \wedge xfz \Rightarrow y = z$".

The convention that f, g, h stand for functions and the notation $f : A \to B$ allow us to be terse (and ungrammatical) when we want : "Let $f : A \to B$ such that ..." means "Let f be a function from A to B, such that ..."

Needless to *emphasise* that f, g, h are generic. We may, and do, use specific symbols for specific functions such as cos, $+$, 1_A.

The terminology *left field, right field, domain, range, onto, total, nontotal, partial, inverse (f^{-1}, as a relation)* and the corresponding definitions are inherited from those for relations (Part IV) and need no further comment. Except one: Note that GS use "partial" to mean "nontotal". This is in conflict with the literature on, for example, computability.

There are two new pieces of terminology for functions

Definition 3.2. A function $f : A \to B$ is 1-1 (algebraists also say *injective*[8]) iff

$$(\forall x)(\forall y)(\forall z)(yfx \wedge zfx \Rightarrow y = z) \tag{1}$$

[7] No loss of generality here: The "in" part could be an n-tuple. Then the in/out *pair* is an $n + 1$-tuple.

[8] Algebraists call "onto" *surjective*.

That is, the function maps distinct inputs to distinct outputs, since (1) says that any two inputs (y and z) that map to the same output (x) must be equal.

A function $f : A \to B$ is a 1-1 correspondence[9] iff it is all three: Total, onto, and 1-1. ■

You must have noticed that neither the definition of function, nor the one of 1-1ness depends on the fields. However, the definition of 1-1 correspondence does depend on both left and right fields.

Remark 3.3. If we rewrite (1) (using sWLUS) in the *equivalent form*

$$(\forall x)(\forall y)(\forall z)(xf^{-1}y \wedge xf^{-1}z \Rightarrow y = z) \tag{1}$$

we have the very important:

$f : A \to B$ is 1-1 iff the *inverse relation* $f^{-1} : B \to A$ is a function.

Example 3.4. $\{\langle 1,2\rangle, \langle 1,3\rangle\}$ is not a function. $\{\langle 1,2\rangle, \langle 2,2\rangle\}$ is a function, but it is not 1-1. $1_A : A \to A$ is a 1-1 correspondence. $\emptyset$ is 1-1 function. ■

Definition 3.5. If $f : A \to B$ is a function, then the formula afb is normally denoted by $f(a) = b$, which is the same as $b = f(a)$. ■

Since functions are relations we can compose them. So, if we have

$$A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D$$

then we can write unambiguously (but informally)

$$A \xrightarrow{f \circ g \circ h} D$$

for their composition *as relations*, without any brackets, because of associativity.

Functions have a peculiar *additional* notation for composition. It is arrived at as follows: Suppose $af \circ gb$. Then, on one hand we have

$$(f \circ g)(a) = b \tag{1}$$

by 3.5. On the other hand, there is a c such that $afcgb$, hence $f(a) = c$ and $g(c) = b$. Substituting c by $f(a)$ in the last one we get

[9] Algebraists also say *bijective*.

$$g(f(a)) = b \tag{2}$$

Comparing (1) and (2) we note an awkwardness: First, there is an order reversal between f and g. Secondly, (2) is more "natural", as it places the input a near the function that will work on it, f. By contrast, (1) places the input next to the function, g, that wouldn't care less about a.

We fix this by adding notation for "function composition" (or "functional composition"):

Definition 3.6 (Functional composition). For functions f and g we define

$$g \bullet f \stackrel{\text{Def.}}{=} f \circ g$$

■

We can now rewrite (1) as

$$(g \bullet f)(a) = b \tag{1$'$}$$

or, combining (1′) and (2), with the implicit understanding that a causes some output (b),

$$(g \bullet f)(a) = g(f(a))$$

And this now looks "natural".

We next turn to inverses.

Definition 3.7 (One-sided inverses). Let us have

$$A \xrightarrow{f} B \xrightarrow{g} A$$

and assume that we have $f \circ g = 1_A$.

Write this $g \bullet f = 1_A$. We say that g is $\underline{\text{a}}$ left inverse of f and f is $\underline{\text{a}}$ right inverse of g. ■

In definition 3.7 note two things:

(1) The emphasis on the indefinite article. One-sided inverses are not unique (see the example that follows).

(2) Who's on left and who's on right is *with respect to functional composition notation*, $\bullet$.

Example 3.8. Let $A = \{a, b\}$, where $a \neq b$, and $B = \{1, 2, 3, 4\}$. Consider the following functions:

$$\begin{aligned}
f_1 &= \{\langle a, 1\rangle, \langle b, 3\rangle\}\\
f_2 &= \{\langle a, 1\rangle, \langle b, 4\rangle\}\\
g_1 &= \{\langle 1, a\rangle, \langle 3, b\rangle, \langle 4, b\rangle\}\\
g_2 &= \{\langle 1, a\rangle, \langle 2, b\rangle, \langle 3, b\rangle\}\\
g_3 &= \{\langle 1, a\rangle, \langle 2, b\rangle, \langle 3, b\rangle, \langle 4, b\rangle\}\\
g_4 &= \{\langle 1, a\rangle, \langle 2, a\rangle, \langle 3, b\rangle, \langle 4, b\rangle\}\\
g_5 &= \{\langle 1, a\rangle, \langle 3, b\rangle\}
\end{aligned}$$

We observe that

$$g_1 \bullet f_1 = g_2 \bullet f_1 = g_3 \bullet f_1 = g_4 \bullet f_1 = g_5 \bullet f_1 = g_1 \bullet f_2 = g_3 \bullet f_2 = g_4 \bullet f_2 = 1_A$$

What emerges is:

(1) The "equation" $x \bullet f = 1_A$ does *not* necessarily have *unique* x-solutions, not even when only *total* solutions are sought.

(2) The equation $x \bullet f = 1_A$ *can* have *nontotal* x-solutions. Neither a total nor a nontotal solution is 1-1 necessarily.

(3) An x-solution to $x \bullet f = 1_A$ can be 1-1 *without being total.*

(4) The equation $g \bullet x = 1_A$ does *not* necessarily have *unique* x-solutions. Solutions do *not* have to be *onto.* ■

In the previous example we saw what we *cannot* infer about f and g from $g \circ f = 1_A$. Let us next see what we *can* infer.

Proposition 3.9. *Given $f : A \to B$ and $g : B \to A$ such that $g \bullet f = 1_A$. Then*

(1) f *is* total *and* 1-1.
(2) g *is* onto.

Proof. (1) Since $g \bullet f$ is total, it follows that f is too: Indeed, I need show that for any $a \in A$ a $b \in B$ exists so that afb. Well, starting from $a1_Aa$ I get $af \circ ga$. Thus, for some $b \in B$, $afbga$. Look no further; we got afb.

Next, let $afc \wedge bfc$, and thus $f(a) = f(b)$. Then $g(f(a)) = g(f(b))$, hence $(g \bullet f)(a) = (g \bullet f)(b)$, that is, $1_A(a) = 1_A(b)$.

Hence $a = b$.

(2) For ontoness of g we argue that there exists an x-solution of the equation $g(x) = a$ for *any* $a \in A$. Indeed, $x = f(a)$ is a solution. ■

Was the onto case too fast? Well, "$g : B \to A$ is onto" means by definition (check Part IV): $\mathrm{ran}(g) = A$, that is

$$\{y|(\exists x \in B)xgy\} = A$$

By the extensionality theorem and writing $g(x) = y$ for xgy I have

$$(\forall y)\Big((\exists x \in B)g(x) = y \equiv y \in A\Big) \tag{1}$$

For a $g : B \to A$ the $\Rightarrow$ direction of (1) is for free (true), hence (1) amounts to (equivalent: By $\mathit{true} \wedge A \equiv A$ and sWLUS)

$$(\forall y)\Big(y \in A \Rightarrow (\exists x \in B)g(x) = y\Big)$$

or

$$(\forall y \in A)(\exists x \in B)g(x) = y$$

In words: For every $y \in A$, I can "solve" $g(x) = y$ for x".

Corollary 3.10. *Not all functions $f : A \to B$ have left (or, right) inverses.*

Proof. Not all functions $f : A \to B$ are 1-1 (or, onto). ■

Corollary 3.11. *Functions with neither left nor right inverses exist.*

Proof. Any $f : A \to B$ which is neither 1-1 nor onto fits the bill. For example, take $f = \{\langle 1, 2\rangle, \langle 2, 2\rangle\}$ from $\{1, 2\}$ to $\{1, 2\}$. ■

Proposition 3.12. *If $f : A \to B$ is a 1-1 correspondence, then $x \bullet f = 1_A$ and $f \bullet x = 1_B$ have the* unique *common solution f^{-1}.*

NB. This unique common solution, f^{-1}, is called *the inverse function* of f. Of course, f^{-1} is the same as the inverse relation of f, but it has additional properties. For starters, it is a function.

Proof. (1) First off, we already know that f^{-1} is a function by remark 3.3. You are also asked to verify that it *is* a common solution in problem set #5.

So we turn to

(2) (Uniqueness of solution) Let $x \bullet f = 1_A$.

Then $(x \bullet f) \bullet f^{-1} = 1_A \bullet f^{-1} = f^{-1}$.[10] By associativity of $\bullet$, this says $x \bullet (f \bullet f^{-1}) = f^{-1}$, i.e., $x = x \bullet 1_B = f^{-1}$. Therefore a left inverse *has* to be f^{-1}. The same can be similarly shown for the right inverse. ∎

Corollary 3.13. *If $f : A \to B$ has both left and right inverses, then it is a 1-1 correspondence, and hence the two inverses equal f^{-1}.*

Proof. From $h \bullet f = 1_A$ (h is some left inverse) follows that f is 1-1 and total. From $f \bullet g = 1_B$ (g is some right inverse) follows that f is onto. ∎

Theorem 3.14 (Algebraic characterization of 1-1ness). *$f : A \to B$ is* total *and* 1-1 *iff it is* left-invertible.[11]

Proof. The *if*-part is proposition 3.9(1). As for the *only if*-part note that $f^{-1} : B \to A$ is single-valued (f is 1-1) and verify that $f^{-1} \bullet f = 1_A$:

For the $\supseteq$ direction pick any $a \in A$ and prove $af \circ f^{-1}a$. Well, f is total, so there is a (unique) b such that afb. Since this is the same as $bf^{-1}a$ we have $afbf^{-1}a$, hence $af \circ f^{-1}a$.

For the $\subseteq$ direction assume that $af \circ f^{-1}b$ and prove that $a = b$, from which $a1_Ab$ follows.

OK, there must be a c such that $afcf^{-1}b$. Hence afc and bfc. By 1-1ness I get $a = b$ as needed. ∎

One can also prove that if $f : A \to B$ is onto, then it is right-invertible, that is, a $g : B \to A$ exists such that $f \bullet g = 1_B$. This result *needs* a new axiom, the axiom of choice so that one can be allowed to pick a potentially infinite set of elements in the sets $\{x \in A | f(x) = y\}$—one element for each $y \in B$.

The text

(1) sweeps the need for the Axiom of Choice under the rug.

[10] The last equality is by the similar result that we proved for relations

$$A \xrightarrow{R} A \xrightarrow{1_A} A$$

The proof when we have

$$B \xrightarrow{f^{-1}} A \xrightarrow{1_A} A$$

where, possibly, $A \neq B$ is identical. You are encouraged to try it out!

[11] That is, it has a left inverse.

(2) (incorrectly) argues as if every set $\{x \in A | f(x) = y\}$, for each $y \in B$ can be enumerated as a sequence with natural number subscripts. This is false in general. E.g., if one of these sets is $\mathbb{R}$—the set of reals—then no such enumeration is possible.

The right thing to do, without messing with a new and esoteric axiom is to leave this story untold.