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

1 Preamble

This section reminds the reader of a couple of important logical tools, and introduces some expositional notation.

Let us recall *generalisation*, first of all. It is the rule:

$$\text{If } \Gamma \vdash A, \text{ and } x \text{ is not free in } \textit{any} \text{ formula in } \Gamma, \text{ then } \Gamma \vdash (\forall x)A \qquad (gen)$$

Now, when Γ is the set of axioms of a "real-life theory",[1] like set theory (about which I write this note) or number theory (Peano arithmetic), at first sight, the restriction on x in (gen) is a bit worrisome.

We worry as follows: Both set theory and Peano arithmetic have an infinite set of axioms. How can we ensure the restriction *in practice*? It is one thing to say—in all those other cases that we saw in class where Γ was finite—that we can always choose in the place of x a variable that has not appeared free in Γ. After all we have a supply of infinitely many variables. But when Γ is infinite itself, for all I know we may have *used* as free *all the variables of our supply* in the formulas of Γ, and as a result no "fresh" variable exists to use in (gen). *Does this mean that in set theory and Peano arithmetic we must disallow generalisation?*

Enough worrying! No. The simple trick is to present all the special axioms, and axiom schemata, in each theory in "closed form", i.e., *without free variables.*

This is always doable! The trick is to eliminate every free variable x in an axiom or axiom schema by placing a $(\forall x)$ in front of the axiom (schema). Intuitively, the axiom (schema) still "says" exactly the same thing as it did before we added the $(\forall x)$.

Definition 1.1. A formula A is called *closed* or a *sentence* iff it has no free variables. ■

Thus, (gen) yields the very usable and important special case:

$$\text{If } \Gamma \vdash A, \text{ and } \textit{all} \text{ formulas in } \Gamma \text{ are closed, then } \Gamma \vdash (\forall x)A \qquad (gen')$$

[1] I mean one you study because it is interesting and applicable, not because it exercises your understanding of MATH1090.

Next let us recall WLUS:

$$\text{If } \vdash A \equiv B, \text{ then } \vdash C[p \setminus A] \equiv C[p \setminus B] \qquad (WLUS)$$

Its usability is restricted when we argue in a theory with non-empty Γ because of the restriction that the hypothesis, $A \equiv B$, be an absolute theorem.

But do not despair! Super-WLUS ("sWLUS") to the rescue. This is

$$\text{If } \Gamma \vdash A \equiv B, \text{ and } \Gamma \text{ contains only } \textit{sentences}, \text{ then } \Gamma \vdash C[p \setminus A] \equiv C[p \setminus B] \qquad (sWLUS)$$

Actually, I'll be asking you to use WLUS to prove rule sWLUS in a forthcoming problem set.

Recall the technique of *Proof By Cases*: If I know that $\Gamma \vdash A_1 \vee \ldots \vee A_n$ and also that $\Gamma \vdash A_i \Rightarrow B$ for $i = 1, \ldots, n$, then I can conclude (by trivial tautological implication) that

$$\Gamma \vdash B$$

In practice, once a proof of $\Gamma \vdash A_1 \vee \ldots \vee A_n$ has been obtained, I do each of $\Gamma \vdash A_i \Rightarrow B$ using the deduction theorem, that is, I add the assumption A_i to Γ and then prove just B.

There is special jargon and organisation for this process.

(i) I prove $A_1 \vee \ldots \vee A_n$
(ii) For $i = 1, \ldots, n$, I say:
Case A_i (this is synonymous to "assume A_i"): Now I do chugga-chugga (one way or another) and I prove B
(iii) Done.

If $\Gamma = \{A_1 \vee \ldots \vee A_n\}$, then the described technique proves $A_1 \vee \ldots \vee A_n \Rightarrow B$.

Here is another (fairly standard) piece of notation: Say A is a formula. We often indicate our *interest* in a variable x that *may* occur in A free by writing either $A[x]$ or $A(x)$.

Just because we are *interested* in x is not any reason to assume that *really* x occurs free in A. This is like writing in elementary math $f(x) = 3$ to define the constant function "3". Even though we wrote "(x)", the function output does not depend on x—x does not occur free in "f".

This type of notation entails the following: Just as $f(x)[x := 9]$ has the shorthand $f(9)$, $A[x][x := t]$ and $A(x)[x := t]$ have the shorthand $A[t]$ and $A(t)$—of course before we employ this kind of shorthand we must be sure that the substitution is defined.

Let me conclude this section by explaining an expositional device: First off, one learns in "Instructor 101" that *in MATH, one must always present the truth, but not necessarily the whole truth.* The reason is obvious: Sometimes the whole truth is bogged down in technical details to the extent that one risks losing the forest for the trees. Thus, I'll offer the "whole truth" in passages that can be skipped on first reading (or even for good). They are identified by at their two ends.

2 Three axioms

Set theory was invented by Georg Cantor in an informal setting. This is what (somewhat unkindly) people nowadays call "Naïve Set Theory".

Because of contradictions that were built-in into the informal approach, mathematicians eventually cleaned up the theory by basing it on a few axiom schemata. Depending on the main contributors' names in a particular axiomatic approach (yes, there are more than one different approaches!) on uses a different acronym for set theory's "Γ", the collection of its special (nonlogical) axioms. For example, the most widely used axiomatization is due to Zermelo and Fraenkel. Its Γ goes under the acronym "ZF" from the contributors' initials.

Neither in the informal theory, nor in the axiomatic theory one defines the notion "set".

In the former one just shrugs the issue off and says that "set" means "collection".[2] In the latter one does not bother at all: One gets to learn what sets are by learning how they behave—i.e., learning theorems about sets. Compare with the similar situation with real numbers: Did anyone ever tell you what $\sqrt{2}$ *really* is? No. Of course, you learned that $(\sqrt{2})^2 = 2$ and that $\sqrt{2}$ obeys the "usual rules of arithmetic and order" and that was more or less all you needed to know.

The one and only fundamental symbol you need in order to do set theory is one predicate of arity 2, namely, $\in$. $S \in T$ is meant to say "S is a member (or element) of T", or "S belongs to T". Using $\in$ as a starting

[2] So, what is a "collection"? I guess, it is a "set".

point one can logically define every other special symbol that people use in set theory.

In this course we will use only a few of those axioms to keep matters simple and relevant to what we want to do. We will call this collection of axioms—our Γ—"ST" (for Set Theory).

In the text's Ch.11 only *one* axiom is really introduced, *Extensionality*.[3] This is

$$(\forall S)(\forall T)(\forall x)\big((x \in S \equiv x \in T) \Rightarrow S = T\big) \tag{Ext}$$

where S, T are <u>set variables</u> (i.e., of type "set") while x is an arbitrary variable (typeless: "set" or "atom" type). Recall that the $\Leftarrow$ direction is not part of the axiom, so it is not included above. In class we proved that direction *logically* (with no nonlogical axioms). Indeed, all we had to note is that

$$(\forall S)(\forall T)(\forall x)\big((x \in S \equiv x \in T) \Leftarrow S = T\big) \tag{1}$$

is an instance of **Ax6** where the syntactic variable "A" in the axiom is here the specific formula "$x \in z$", and we do the substitutions $[z := S]$ and $[z := T]$. By the distributivity of $\forall$ over $\wedge$, (Ext) and (1) give the *Extensionality Theorem*:

$$(\forall S)(\forall T)(\forall x)\big((x \in S \equiv x \in T) \equiv S = T\big) \tag{2}$$

Note how our first axiom is indeed *closed* (a sentence—this connects to the discussion in section 1).

Example 2.1. By extensionality, $\{1, 2\} = \{2, 1, 1, 2\}$ since every x in the lhs is in the rhs and vice versa. In particular, extensionality implies that order of appearance and multiplicity of appearance of set elements is irrelevant. ■

Disclaimer. It is beyond our scope—and beyond our goals—to present set theory in a strict axiomatic setting. We will adopt a few, *but nowhere near all*, axioms and will draw useful consequences from them, logically. Our planned omissions will force us to do a "quasi-formal" set theory, that is, one where we will assume *parts of mathematics outside set theory.* The formal part will be served by giving (set theory) axioms, definitions, and logical proofs of a number of theorems.

[3] Everything else the authors call an "axiom" is really a definition.

The informal part will consist of those instances where we will bring in examples (say from the integers, reals, etc.), but most significantly where we will pretend that we can have set theory and other theories (about numbers and other mathematical objects) interact and coexist.

Specific example from the text: There is an axiom about the number of elements in a set. The way it is given, the authors *tacitly* pretend

(1) that we know enough about the natural numbers to know what "$\sum_{x \in S} 1$" means, that is, how sums like "$\sum_{1 \leq x \leq n} f(x)$" are *formally* defined; what their properties are; etc. That is too much to pretend about because—up to the point where the axiom was given—not a single axiom about the natural numbers was presented.

(2) that we can mix set theory and other theories. While (2) is, in principle, possible it is never done *formally*; it is very messy.

So what will we do? Well, we will first of all not be totally formal! Secondly, we will mix, in our quasi-informal/quasi-formal manner, results from theories that *we pretend to know.* This is in the spirit of (1)–(2) above. However, for the sake of avoiding to prove falsehoods or present circular proofs, we will avoid examples such as (1), where we have to pretend that we know much more than we really have the right to pretend that we know. In short, we will use a *minimal amount* of "mathematics" outside set theory (In particular, we will use a different definition regarding size of sets.)

<u>Fact</u>: One can base all mathematics on set theory. That is, one can base all mathematics on the axioms of set theory without introducing any additional axioms for other mathematical theories. In particular, one can even "build" the natural numbers and their properties within set theory, then the integers, then the rationals, then the reals, etc. *But we will not take this approach as it is far too advanced for our purposes and scope.*

Next we need some definitions so we can work with the symbol "$\{x|A\}$" that is known as a *set term.*

Intuitively, this symbol denotes the collection of all those values of x that make A true. In other words, if S is a set variable and we state "$S = \{x|A\}$", what we mean by that is

(For all possible values x) x is in S iff it passes the "entrance test" A.

This motivates:

Definition 2.2. If S is a set variable, then the "text" $S = \{x|A\}$ (also written as $S = \{x : A\}$) is an abbreviation for

$$(\forall x)(x \in S \equiv A)$$

$\blacksquare$

Thus, the tautology $(\forall x)(x \in S \equiv A) \equiv (\forall x)(x \in S \equiv A)$ leads to the absolute theorem (indeed tautology)

$$S = \{x|A\} \equiv (\forall x)(x \in S \equiv A)$$

Now, "$S = \{x|A\}$" where S is a set variable says "$\{x|A\}$ is a set named S".

But what else could a collection like $\{x|A\}$—the collection of all x that make A true—be? Well, it might be a collection that is "too big" or "too complex" to be a set.

For example, we have seen that for some choices of A we are in trouble: Russell has shown (see my additional Hilbert proofs examples on the MATH2090/Fall 03 URL) that we can prove

$$\neg(\exists y)(\forall x)(x \in y \equiv \neg x \in x) \tag{1}$$

in pure logic. Thus, "let $S = \{x|x \notin x\}$" is wishful thinking: If this were possible (non contradictory) we would be able to obtain—by $\exists$ introduction—the negation of (1) above; and that is impossible.

Compare: If y is a real number variable, we know what $y = \sqrt{-1}$ means. It means $y^2 = -1$. We also know that this is impossible. Thus, the questions of what an abbreviation means and whether it is valid (or provable) are two different issues.

Just as (for y a real variable and t a term) "let $y = t$" says "assume t is a real y", in the same way, if S is a set variable, "let $S = \{x|A\}$" says "assume $\{x|A\}$ is a set S".

Reference to y (respectively, S) is eliminated if we state instead $(\exists y)(y = t)$ (respectively, $(\exists S)(S = \{x|A\})$). For example, "$(\exists S)(S = \{x|A\})$" or

$$(\exists S)(\forall x)(x \in S \equiv A)$$

say "$\{x|A\}$ is a set".

We next translate the text

$$x \in \{x|A\} \tag{2}$$

if $\{x|A\}$ *represents a set.*

Theorem 2.3 ($\in$-elimination). *If $\{x|A\}$ is a set, then* (2) *above is provably (in ST) equivalent to A.*

Proof. We give two different proofs.

(1) Informal proof (like the ones a mathematician or a computer scientist might construct):

Let $S = \{x|A\}$ where S has type set. This is a legitimate assumption since $\{x|A\}$ is a set.

Thus by 2.2 $(\forall x)(x \in S \equiv A)$ and hence, by specialisation,

$$x \in S \equiv A$$

Replacing S by $\{x|A\}$ we are done.

(2) Formal proof.

(1)	$(\exists y)(y = \{x|A\})$	(assume)
(2)	$S = \{x|A\}$	(by (1): assume; S fresh)
(3)	$(\forall x)(x \in S \equiv A)$	((2) and 2.2)
(4)	$x \in S \equiv A$	((3) and spec.)
(5)	$x \in \{x|A\} \equiv A$	((2,4), **Ax6** and taut. impl.)

■

Note that the formal proof faithfully tracks the informal one. For example, step (2) is the part "Let $S = \{x|A\}$" in the informal one. You are encouraged to design informal proofs before you plunge into chugga-chugga. This is like programming in pseudo-code prior to actual "coding". It just makes things clearer.

By the way, since the "assumption" (1) was presumably obtained using properties (axioms) of sets, that is why this is a set theory theorem rather than an absolute theorem.

Corollary 2.4. *If z is fresh and if $\{x|A[x]\}$ is a set, then so is $\{z|A[z]\}$ and moreover $\{x|A[x]\} = \{z|A[z]\}$.*

Proof. I give an informal proof only: Let $S = \{x|A[x]\}$ where S is of type set. Thus (by 2.2),

$$(\forall x)(x \in S \equiv A[x]) \tag{1}$$

Using dummy renaming in (1) we get

$$(\forall z)(z \in S \equiv A[z]) \tag{2}$$

By 2.2 and (2) we have proved $S = \{z|A[z]\}$. In particular, $\{z|A[z]\}$ is a set. Transitivity of "=" yields the last claim of the corollary. ■

In particular we get three important results:

(i) x is a "dummy" in $\{x|A[x]\}$ under the corollary's assumptions. Indeed, $\{z|A[z]\}$ has no occurrence of x at all, and $\{x|A[x]\} = \{z|A[z]\}$.

(ii) With all the assumptions as in the corollary, $z \in \{x|A[x]\}$ is provably equivalent (in theory ST) to $A[z]$, since it is the same as $z \in \{z|A[z]\}$ (and we invoke 2.3).

(iii) Indeed, (ii) above can be strengthened by dropping the requirement that z is "fresh". Let t be *any term* (such as variable, constant, function application) for which $A[x := t]$ is defined. Then I can prove

$$ST \vdash t \in \{x|A\} \equiv A[x := t] \tag{3}$$

provided we know that $\{x|A\}$ is a set. I can write (3) also as

$$ST \vdash t \in \{x|A[x]\} \equiv A[t] \tag{3$'$}$$

The proof of (3), whether formal or informal, is exactly as in 2.3: Proceeding informally, let S be a set-type variable not appearing in the term t, and write $S = \{x|A\}$—legitimate since $\{x|A\}$ is a set. By 2.2, I have $(\forall x)(x \in S \equiv A)$. Specialisation yields

$$(x \in S \equiv A)[x := t]$$

that is,

$$t \in S \equiv A[x := t]$$

Replacing S by its equal (using Ax6) I finally have proved (3). Note that S not appearing in t is not a serious restriction, since I have a supply of infinitely many variables to choose from. It ensures that when I plug "$\{x|A\}$" into S in "$t \in S$" this does not also get plugged into t.

Recall the definition from class:

Definition 2.5. The "text" $\{t[\vec{x}] : A[\vec{x}]\}$ stands for $\Big\{z : (\exists \vec{x})\big(t[\vec{x}] = z \wedge A[\vec{x}]\big)\Big\}$ where "$\vec{x}$" is short for $x_1, x_2, \ldots, x_n$ and is here the list of *active* or *linking* or *dummy* variables, and z is fresh. ■

In the text, GS indicate the linking variables like this

$$\{\underbrace{\vec{x} : \text{type}}_{\text{links}} | A : t\}$$

Following standard notation, we indicate them instead by a comment, or by using $[\vec{x}]$ next to A (the "property") and to t (the generic term we are collecting).

Example 2.6. Consider $S = \{x + y | 0 < x < y\}$ where x, y are of type $\mathbb{N}$ (natural number).

(1) Suppose x only is active. Eliminate it to get $S = \{y + 1, y + 2, \ldots, 2y - 1\}$.

We always eliminate the active variables. The non-active variables are called parameters. They are *constants* that we have not disclosed their values.

(2) Suppose y only is active. Eliminate it to get $S = \{2x + 1, 2x + 2, \ldots\}$.

(3) Suppose both x and y are active. Eliminate them to get $S = \{3, 4, 5, \ldots\}$. ■

The following is very important.

Theorem 2.7. *If* $\{x : A\}$ *and* $\{x : B\}$ *are sets, then*

$$ST \vdash \{x : A\} = \{x : B\} \equiv (\forall x)(A \equiv B) \tag{$*$}$$

Proof.

$$\begin{aligned}
&\{x : A\} = \{x : B\} \\
\equiv &\Big\langle \text{Ext. theorem \& assumption} \Big\rangle \\
&(\forall x)(x \in \{x : A\} \equiv x \in \{x : B\}) \\
\equiv &\Big\langle \text{sWLUS (twice) \& } \in\text{-elim} \Big\rangle \\
&(\forall x)(A \equiv B)
\end{aligned}$$

■

We introduce some interesting collections:

Definition 2.8. $\{x|\mathit{false}\}$ is denoted by $\emptyset$ and is pronounced "the empty (or void, or null) set".

$\{x : \mathit{true}\}$ is denoted by $\mathbb{V}$ and is pronounced "the (absolute) universe". ∎

The nomenclature is obviously motivated: Nothing passes the entrance test in $\emptyset$, whereas everything passes the entrance test in $\mathbb{V}$.

It turns out that $\emptyset$ is a set. While this can be proved from additional assumptions we will take the easy way out:

Axiom 2.9. $\{x|\mathit{false}\}$ is a set. ∎

Since "$\emptyset$" denotes a fixed set it is a constant, of course, of type set.

On the other hand, we will eventually see that $\mathbb{V}$ is too big and too complex to be a set. Like $\{x|x \notin x\}$, it is a non-set collection. Set theorists call such collections *proper classes*.

Proposition 2.10. $ST \vdash S = \emptyset \equiv (\forall x)x \notin S$.

Proof.

$$
\begin{aligned}
& S = \emptyset \\
\equiv & \left\langle \text{Ext. theorem} \right\rangle \\
& (\forall x)(x \in S \equiv x \in \{x : \mathit{false}\}) \\
\equiv & \left\langle \text{sWLUS \& } \in\text{-elim} \right\rangle \\
& (\forall x)(x \in S \equiv \mathit{false}) \\
\equiv & \left\langle \text{WLUS} \right\rangle \\
& (\forall x)x \notin S
\end{aligned}
$$

∎

We have at once

Corollary 2.11. $ST \vdash S \neq \emptyset \equiv (\exists x)x \in S$.

Definition 2.12. For any set variables S and T we write $S \subseteq T$—pronounced "S is a subset of T" (also: "T is a superset of S")—as a shorthand for $(\forall x)(x \in S \Rightarrow x \in T)$.

We write $S \subset T$—pronounced "S is a *proper* subset of T" (also: "T is a proper superset of S")—as a shorthand for $S \subseteq T \wedge S \neq T$. ■

Joining $(\forall x)(x \in S \Rightarrow x \in T)$ and $(\forall x)(x \in S \Leftarrow x \in T)$ via the $\forall - \wedge$ distributivity theorem we have at once (via the Extensionality theorem)

Proposition 2.13. $ST \vdash S \subseteq T \wedge T \subseteq S \equiv S = T$.

Clearly, the $\Rightarrow$ is contributed by the Extensionality axiom, while $\Leftarrow$ is contributed by **Ax6**. The proposition is a useful tool to prove two sets equal. One proves the $\subseteq$ and $\supseteq$ directions separately.

We also have a counterpart to 2.7

Proposition 2.14. *If* $\{x : A\}$ *and* $\{x : B\}$ *are sets, then*

$$ST \vdash \{x : A\} \subseteq \{x : B\} \equiv (\forall x)(A \Rightarrow B) \qquad (*)$$

Proof.

$$\begin{aligned}
&\{x : A\} \subseteq \{x : B\} \\
\equiv\ &\Big\langle \text{definition 2.12} \Big\rangle \\
&(\forall x)(x \in \{x : A\} \Rightarrow x \in \{x : B\}) \\
\equiv\ &\Big\langle \text{sWLUS (twice) \& } \in\text{-elim} \Big\rangle \\
&(\forall x)(A \Rightarrow B)
\end{aligned}$$

■

Example 2.15. For any set S,

(1) $\emptyset \subseteq S$

(2) $S \subseteq S$.

Well, for (1) I need (see 2.12, 2.9 and $\in$-elimination) to prove $(\forall x)(\mathit{false} \Rightarrow A)$ which is in **Ax1**.

For (2) I need to prove $(\forall x)(x \in S \Rightarrow x \in S)$. This is also in **Ax1**.

Pause. Which of the two, (1) and (2) is an absolute theorem? Why? ■

Here are two easy theorems:

Theorem 2.16. $ST \vdash \emptyset \neq S \subseteq T \Rightarrow T \neq \emptyset$.

Proof. Informal proof. Obviously, by the deduction theorem, we prove instead $\overline{ST, \emptyset \neq S}, S \subseteq T \vdash T \neq \emptyset$. By the 1st assumption, let $z \in S$. By the 2nd assumption and definition of "$\subseteq$" I get $z \in T$. Hence $T \neq \emptyset$.

Formal proof. This faithfully tracks the above.

(1)	$(\exists x)x \in S$	$\big($assume$\big)$
(2)	$z \in S$	$\big($by (1): assume; z is fresh$\big)$
(3)	$(\forall x)(x \in S \Rightarrow x \in T)$	$\big($assume$\big)$
(4)	$z \in S \Rightarrow z \in T$	$\big($(3) & spec.$\big)$
(5)	$z \in T$	$\big($(2,4) & MP$\big)$
(6)	$(\exists x)x \in T$	$\big($(5) & $A[x := t] \vdash (\exists x)A$ rule$\big)$

■

Theorem 2.17.

$$S \subset T \Rightarrow T \neq \emptyset$$

Proof. The informal proof is very easy: Since $S \subset T$, there is *something* in T—call it z—that is not in S (otherwise, $T \subseteq S$ which along with the "$S \subseteq T$" part of the hypothesis yields $S = T$. This contradicts half the hypothesis). But $z \in T$ means $T \neq \emptyset$.

The formal proof below does not track the above. Invoke the deduction theorem, and assume hypothesis (line (1) below).

(1)	$(\forall x)(x \in S \Rightarrow x \in T) \wedge \neg(\forall x)(x \in S \equiv x \in T)$	$\big($assume$\big)$
(2)	$(\forall x)(x \in S \Rightarrow x \in T)$	$\big($(1) and Post's theorem$\big)$
(3)	$\neg(\forall x)(x \in S \equiv x \in T)$	$\big($(1) and Post's theorem$\big)$
(4)	$(\exists x)(x \in S \equiv \neg x \in T)$	$\big($(3), Post and WLUS$\big)$
(5)	$z \in S \equiv \neg z \in T$	$\big($by (4): <u>assume</u>; z fresh$\big)$
(6)	$z \in S \Rightarrow z \in T$	$\big($(2) and specialisation$\big)$
(7)	$z \in T$	$\big($(5,6) and Post$\big)$
(8)	$z \in T \Rightarrow (\exists x)x \in T$	$\big($$\exists$-introduction theorem$\big)$

(9) $(\exists x)x \in T$ $\Big((7,8) \text{ and MP}\Big)$

But the last line says "$T \neq \emptyset$". ■

We have already seen that the collection $\{x|x \notin x\}$ is not a set. Thus, unlike what the philosopher Frege suggested, it is *not* the case that all collections $\{x|A\}$ are sets. Our second axiom said that if $A \equiv$ *false* then $\{x|A\}$ *is* a set.

Are there general conditions on A that guarantee the set-term $\{x|A\}$ to be a set?

Yes, and what these conditions do is to ensure—*intuitively speaking*—that the resulting collection is not too big or too complex. The easiest way is to tweak A with an additional condition, as in the axiom below:

Axiom 2.18 (Axiom of subsets). For any set variable T and any formula A of set theory, $\{x|A \wedge x \in T\}$ is a set. ■

As we already know how to translate from English to symbolic language, the axiom says, without words:

$$(\forall T)(\exists y)(\forall x)(x \in y \equiv A \wedge x \in T) \tag{1}$$

Note that the arbitrary condition A was modified into $A \wedge x \in T$.

First off, why "axiom of subsets"?[4] Well, if I let $S = \{x|A \wedge x \in T\}$ where S is a set variable (a legitimate action according to the axiom) I can then prove $S \subseteq T$. Indeed,

$$\begin{aligned}
& S \subseteq T \\
\equiv & \Big\langle \text{def. of } \subseteq \Big\rangle \\
& (\forall x)(x \in \{x|A \wedge x \in T\} \Rightarrow x \in T) \\
\equiv & \Big\langle \text{sWLUS \& } \in\text{-elim.} \Big\rangle \\
& (\forall x)(A \wedge x \in T \Rightarrow x \in T)
\end{aligned}$$

But the bottom line is in **Ax1**.

Thus, the axiom says that "*parts* of sets (here T is the given set, S is a part) are small and simple enough to be sets themselves".

[4] It is also called "separation axiom".

Clearly the axiom can be expressed in *closed form* so that the requirement that no axiom in ST has any free variables is met.[5] One applies the usual trick: For any formula A of free variables $x_1, \ldots, x_n$, for any fresh set variable T, the following is an axiom of ST:

$$(\forall T)(\forall x_1) \cdots (\forall x_n)(\exists y)(\forall z)(z \in y \equiv A \wedge z \in T)$$

[5] How about the axiom for $\emptyset$?