Extending Benson Mates’ System of Logic to the Second Order

Phillip Lane

The University of Alabama in Huntsville

Abstract

Second-order logic is an expansion upon first-order logic where quantifiers are permitted to quantify not only over first-order variables, but also over second-order variables, including both predicates and functions. While this brings an entirely new set of sentences that are easily expressed while otherwise impossible to express in a first-order language, it is not without its setbacks. I will attempt to take a formulation of second-order logic, as well as a derivation system, as expressed by Stewart Shapiro in his book, *Foundations without Foundationalism: A Case for Second-Order Logic*, and apply it to Benson Mates' derivation system for $\mathcal{L}'$ in *Mates Elementary Logic*. Different properties will be discussed, including fundamental differences between first-order logic, and similarities to other higher-order logics. Finally, the arguments for and against second-order logic will briefly be mentioned.

Extending Benson Mates' System of Logic to the Second Order

The reader should be familiar with *Mates Elementary Logic* through at least Chapter 10, where Mates discusses Axioms for $\mathcal{L}_{\mathrm{I}}$. We will expand the language $\mathcal{L}'$ by taking the steps Shapiro uses to expand the deductive system $D1$ to $D2$ and applying those to the deductive system for $\mathcal{L}'$.

Before we can begin expanding Mates' system of logic as put forth in *Mates Elementary Logic*, clarifications about notation must be made. Because we are essentially combining two systems of logic by two different authors, there are radical notational differences. These ambiguities must be discussed.

Removing Notational Ambiguity

Firstly, I will be using the turned A to represent universal quantification, where Mates uses no logical constant at all. Secondly, the notation $\langle x\rangle_n$ represents n sequential first-order variables. For example, $(\forall\langle x\rangle_2)P^2\langle x\rangle_2$ is equivalent to the sentence $(\forall x)(\forall y)P^2xy$. This notation is important when talking about quantification over arbitrary-degree predicates and functions. Finally, the notation $\phi(a, x, P^2, f^4, \dots)$ represents any formula ϕ that includes the listed variables and non-logical constants. The variables are assumed to be free. Therefore, $\phi(F^1, x)$ could mean F^1x, or G^2xa & F^1b, or any other formula, given it contains both F^1 and x. It is worth noting that if ϕ appears twice in the same rule, axiom, etc. it is taken to mean the same ϕ. Therefore, $\phi(P) \rightarrow \phi(Q)$ is taken to mean something to the effect of $(\forall x)Px \rightarrow (\forall x)Qx$.

If a notation is not discussed in the previous paragraph, it is safe to assume it is equivalent to how the language $\mathcal{L}'$ defines a well-formed formula.

Crafting Well-Formed Formulas

The first thing we have to do to craft well-formed formulas. Because, later on, we will be dealing with axiom schemas in our derivation system, our well-formed formulas will be a bit broader than what Mates would allow in $\mathcal{L}'$.

First, as Mates discusses grammar of $\mathcal{L}'$, we will discuss grammar of $\mathcal{L}_2$.

1. *Variables*
    a. *First-order variables* are represented by the lower-case letters 'u' through 'z,' possibly having a numerical subscript.
    b. *Second-order predicate variables* are represented by the upper-case letters 'U' through 'Z,' possibly with a numerical superscript denoting its degree, and possibly having a numerical subscript.
    c. *Second-order function variables* are represented by the lower-case letters 'f' through 'h,' with a numerical superscript denoting its degree, and possibly having a numerical subscript.
2. *Constants*
    a. *Logical constants* are any of the following symbols:

    $$- \quad \& \quad \vee \quad (\quad) \quad \rightarrow \quad \leftrightarrow \quad \forall \quad \exists \quad = \quad \neq$$

    b. *Non-logical constants* fall into a few classes:
        i. *First-order constants* are represented by the lower-case letters 'a' through 'e,' or 'i' through 't,' possibly having a numerical subscript.
        ii. *Predicates* or *second-order predicate constants* are represented by the upper-case letters 'A' through 'T,' possibly with a numerical

superscript denoting its degree, and possibly having a numerical subscript.

iii. *Functions* or *second-order function constants* are represented by the lower-case letters 'i' through 't,' with a numerical superscript denoting its degree, and possibly having a numerical subscript.

A *predicate of degree n* is a second-order predicate variable or constant with a positive superscript n denoting its degree.

A *sentential letter* is a predicate without a superscript.

An *individual symbol* is a first-order variable or first-order constant.

Atomic formulas are crafted in any of the following ways:

a. A sentential letter is an atomic formula.
b. If α and β are individual symbols, $\alpha = \beta$ and $\alpha \neq \beta$ are atomic formulas.
c. A predicate of degree n followed by n individual symbols is an atomic formula.

A *formula* is crafted as follows:

a. All atomic formulas are formulas.
b. If ϕ is a formula, so is $-\phi$.
c. If ϕ and ψ are formulas, so are $(\phi \vee \psi)$, $(\phi \,\&\, \psi)$, $(\phi \rightarrow \psi)$, and $(\phi \leftrightarrow \psi)$.
d. If ϕ is a formula and ω is a variable, then $(\forall\omega)\phi$ and $(\exists\omega)\phi$ are formulas.
e. Nothing else is a formula.

Free/bound variables and sentences are defined in the usual manner. We will also allow dropping of outermost parenthesis (given it introduces no ambiguity) and the superscript next to predicates (given it introduces no ambiguity).

Semantics

We will begin by talking about the semantics of a sentence or formula of second-order logic. This is important for identifying valid sentences and being able to translate the natural language into $\mathcal{L}_2$. There are two main types of semantics that we will discuss, the first being *standard semantics* and the second being *Henkin semantics*. There is a third type of semantics called *first-order semantics*, but this is similar enough to Henkin semantics to be uninteresting to us.

In standard semantics, a model M of a formula $\mathcal{L}_2$ is represented by the structure $\langle d, I \rangle$, where d is the domain (or universe of discourse) and I is an interpretation that assigns members of the domain to each non-logical constant. Since $\mathcal{L}_2$ is an expansion upon $\mathcal{L}'$, this means the interpretation must exclude defining any relation for I_1^2, and instead, I_1^2 is the equality relation. The model is accompanied by the variable assignment function s, which assigns to each first-order variable an element of d, to each second-order predicate variable of degree n a subset of d^n (that is, an element of $\mathcal{P}(d^n)$ where $\mathcal{P}$ is the powerset operation), and to each second-order function variable of degree n a mapping from d^n to d, where d^n is the Cartesian product of d with itself $n - 1$ times. Therefore, $d^1 = d$, $d^2 = d \times d$, $d^3 = d \times d \times d$, and so on.

We use the notation $M, s \vDash \phi$ to denote that ϕ is a semantic consequence of a model M and a variable assignment function s. If ϕ is a sentence, we could also say $M \vDash \phi$. The notions of universal quantification, validity, satisfiability, and consequence is defined as follows:

- The sentence $(\forall\omega)\phi$ is true under a model M if and only if $M, s \vDash \phi$, for all s where each s is identical to each other except possibly in their assignment of ω. Note that since s for predicates can span over each element of $\mathcal{P}(d^n)$, that means the universal quantifier must span over each element thereof. Similarly, for functions, the universal quantifier must span over all possible functions in M.
- A sentence ϕ is *standardly valid* if and only if, for all models M, it is true that $M \vDash \phi$.
- A sentence ϕ is *standardly satisfiable* if and only if, for some model M, it is true that $M \vDash \phi$.
- A set of sentences Γ is *standardly satisfiable* if and only if, for some model M, it is true for all sentences ϕ_n in Γ that $M \vDash \phi_n$.
- A sentence ϕ is a *consequence* from a set of sentences Γ if and only if $\Gamma \cup \{-\phi\}$ is not standardly satisfiable.

In Henkin semantics, a model M^H of a formula $\mathcal{L}_2$ is represented by the structure $\langle d, D, F, I\rangle$, where d is the domain (or universe of discourse) and I is an interpretation function that assigns members of the domain to each non-logical constant. D represents the range of all second-order predicate variables, and is defined as $D(n) \subseteq \mathcal{P}(d^n)$. That is, $D(n)$ is a subset of the powerset of d^n. F likewise represents the range of all second-order function variables and is defined as $F(n) \subseteq \{\, f \mid f: d^n \to d \,\}$, that is, $F(n)$ is a subset of all of the functions that map members of d^n to members of d. This is also accompanied by a variable assignment function s (sometimes denoted s^H), which assigns to each first-order variable an element of d, to each second-order predicate variable of degree n an element of $D(n)$, and to each second-order function variable of degree n an element of $F(n)$. We can refer to a *full Henkin model* if and only if $D(n) = \mathcal{P}(d^n)$ and $F(n) = \{\, f \mid f: d^n \to d \,\}$. Such a model is denoted M^F.

We use the notation $M^H, s \vDash \phi$ to denote that ϕ is a semantic consequence of a Henkin model M^H and a variable assignment function s. Likewise, we can say $M^H \vDash \phi$ if ϕ is a sentence. The ideas of validity, satisfiability, and consequence are effectively the same as standard semantics, with a key difference being in the semantics of universal quantification. The definition is the same, but the range of universal quantifiers for predicates is no longer over the set $\mathcal{P}(d^n)$, but rather, over the set D. The range of functions is similarly limited as the set F defines.

The primary difference between standard semantics and Henkin semantics is that standard semantics makes the language properly second-order. Henkin semantics preserves properties such as completeness and compactness, but loses the expressive power that standard semantics gives. For this reason, standard semantics are far more powerful, and are assumed for the rest of this paper.

Expansion of $\mathcal{L}'$'s Derivation System to $\mathcal{L}_2$

Shapiro begins by introducing a few languages, then following up with providing deductive systems for the more substantial ones. He begins with describing the language $L1K$, which is most similar to the Matesian language we are most familiar with, $\mathcal{L}$. He then extends the language to $L1K =$, which is most similar to $\mathcal{L}_I$. A description is then given of a language $L2K$, which is an extension on $L1K$ with second-order variables (including both predicates and function symbols). Note that $L2K$ is *not* an extension of $L1K =$. Finally, he talks about $L2K -$, which is $L2K$ with identity.

We will begin a little differently, though. We will instead begin with the deductive system Mates gave for $\mathcal{L}'$ and expand to get a deductive system for $\mathcal{L}_2$. Shapiro gives a deductive system of $L1K$ and called $D1$, which uses several axioms and two rules of deduction to give a sound, complete system of deduction. Since we know that the rules in $\mathcal{L}'$ are also sound and complete, we can simply expand on $\mathcal{L}'$'s derivation system the same way Shapiro expands on $D1$ to get $D2$. We will be slightly modifying his rules, however, to better fit the trends of how Mates structured $\mathcal{L}'$.

The rules and axioms that Shapiro gives for the expansion of $D1$ to $D2$ are as follows:

1. $(\forall X^n)\phi(X^n) \rightarrow \phi(P^n)$ (universal specification over predicates)
2. $(\forall f^n)\phi(f^n) \rightarrow \phi(p^n)$ (universal specification over functions)
3. $\phi \rightarrow \psi(P^n) \vdash \phi \rightarrow (\forall X^n)\psi(X^n)$ where P^n does not appear in any premise of the deduction (universal generalization over predicates)
4. $\phi \rightarrow \psi(p^n) \vdash \phi \rightarrow (\forall f^n)\psi(f^n)$ where p^n does not appear in any premise of the deduction (universal generalization over functions)
5. $(\exists X^n)(\forall\langle x\rangle_n)\big(X^n\langle x\rangle_n \leftrightarrow \phi(\langle x\rangle_n)\big)$ (axiom schema of comprehension i.e. every formula is representable by a relation, and vice versa)
6. $(\forall X^{n+1})\big((\forall\langle x\rangle_n)(\exists y)X^{n+1}\langle x\rangle_n y \rightarrow (\exists f^n)(\forall\langle x\rangle_n)X^{n+1}\langle x\rangle_n f^n\langle x\rangle_n\big)$ (axiom of choice)

We are going to revise the first four to better fit our purposes. Starting with 1 and 2, by using the corresponding conditional, we will turn these into rules of derivation instead of axioms. This means we instead have:

1. $(\forall X^n)\phi(X^n) \vdash \phi(P^n)$
2. $(\forall f^n)\phi(f^n) \vdash \phi(p^n)$

We will, from here on, reference these rules as rule *US*, as with its first-order counterpart. As for rules 3 and 4, we will generalize these a little bit further by removing the conditional. This action gives us:

3. $\phi(P^n) \vdash (\forall X^n)\phi(X^n)$ where P^n does not appear in any premise of the deduction

4. $\phi(p^n) \vdash (\forall f^n)\phi(f^n)$ where p^n does not appear in any premise of the deduction

Next, we modify the axiom schema of comprehension such that the conditional only flows one way (from function schema to relations) instead of being a biconditional. This makes it far easier to prove soundness about the rule. We add two more rules to generalize rule *E* to the second order. Finally, we will add one more rule of derivation, which will be explained later. Doing this will result in our final set of rules that we will use to extend $\mathcal{L}'$ to $\mathcal{L}_2$:

1. $(\forall X^n)\phi(X^n) \vdash \phi(P^n)$ (universal specification over predicates, or *US*)
2. $(\forall f^n)\phi(f^n) \vdash \phi(p^n)$ (universal specification over functions, or *US*)
3. $\phi(P^n) \vdash (\forall X^n)\phi(X^n)$ where P^n does not appear in any premise of the deduction (universal generalization over predicates, or *UG*)
4. $\phi(p^n) \vdash (\forall f^n)\phi(f^n)$ where p^n does not appear in any premise of the deduction (universal generalization over functions, or *UG*)
5. $-(\forall X^n) - \phi \vdash (\exists X^n)\phi$ and $(\exists X^n)\phi \vdash -(\forall X^n) - \phi$ (quantifier exchange over predicates, or *E*)
6. $-(\forall f^n) - \phi \vdash (\exists f^n)\phi$ and $(\exists f^n)\phi \vdash -(\forall f^n) - \phi$ (quantifier exchange over functions, or *E*)
7. $(\forall X)(Xa \leftrightarrow Xb) \vdash a = b$ (identity of indiscernibles, or *I*)

This ruleset is accompanied by the following two axioms:

I. $(\exists X^n)(\forall\langle x\rangle_n)(\phi(\langle x\rangle_n) \rightarrow X^n\langle x\rangle_n)$ (axiom schema of comprehension)

II. $(\forall X^{n+1})\big((\forall\langle x\rangle_n)(\exists y)X^{n+1}\langle x\rangle_n y \rightarrow (\exists f^n)(\forall\langle x\rangle_n)X^{n+1}\langle x\rangle_n f^n\langle x\rangle_n\big)$ (axiom of choice)

This, along with the rules *P*, *T*, *C*, first-order *US*, first-order *UG*, and first-order *E* complete the rule/axiom set for $\mathcal{L}_2$. We must make clarifications about how the axioms are used. When invoking an axiom, it must be made into a well-formed formula. This is done by specifying at the time of use the degrees of all second-order predicate variables and second-order function variables. One is also required to replace $(\forall\langle x\rangle_n)$ and $(\exists\langle x\rangle_n)$ with a train of n universal quantifiers or existential quantifiers respectively, each quantifying over a single variable, renaming other variables if necessary. Finally, all schemas must be replaced with a well-formed formula. So, for instance, an instantiation of the axiom schema of comprehension might be:

$$(\exists X^2)(\forall x)(\forall y)\big((P^1x \rightarrow P^1y) \rightarrow X^2xy\big)$$

This would be a well-formed variant thereof.

My current research is focusing on formulating equivalent rules of derivation to the axioms, to transform the current system of $\mathcal{L}_2$ into a system of natural deduction. This is done by using the axioms to prove sentences of the form $\phi \rightarrow \psi$, and then using those sentences to prove the axiom. This proves equivalent power between the axioms and the rules.

We will now discuss why the last rule of derivation was added as an expansion on rule *I*. This ties directly into why we began with $\mathcal{L}'$ and expanded to $\mathcal{L}_2$ when, alternatively, we could've started with $\mathcal{L}$ and used Shapiro's definition of equality to form an axiom. He defines equality where $\alpha = \beta$ is taken as shorthand for $(\forall X)(X\alpha \leftrightarrow X\beta)$. This means an equivalent axiom would be the following:

$$(\forall x)(\forall y)\big(x = y \leftrightarrow (\forall X)(Xx \leftrightarrow Xy)\big)$$

This has a rather substantial consequence: identity of indiscernibles. The derivation of this fact is a simple application of rule *US* twice, then rule *T*, then rule *UG* twice:

$$(\forall x)(\forall y)\big((\forall X)(Xx \leftrightarrow Xy) \rightarrow x = y\big)$$

which states that, for any two objects, if all properties apply either to both or neither, then the objects must be identical.

One would expect this to be the reason I chose to expand upon $\mathcal{L}'$ instead of $\mathcal{L}$, but that is not the case. Rather, the choice was a lot simpler: using Mates' method allowed for less axioms. We could have equivalently used Shapiro's axiom of equality, and to prove it, I will first prove identity of indiscernibles under standard semantics for $\mathcal{L}_2$.

Proof: suppose the interpretation defines two first-order constants α and β. Consider the case where if a property applies to α, it applies to β, and vice versa (that is, $(\forall X)(X\alpha \leftrightarrow X\beta)$). In standard semantics, when we discuss all properties of degree 1, we discuss elements of $\mathcal{P}(d)$. Therefore, if any property, when true for one is true for the other, the two first-order constants must denote the same object in the domain, for if they didn't, there would exist some property for which one is true and the other is not. A basic example is this: if α denotes "ball" and β denotes "very similar ball," there must exist a property $P = \{\text{"ball"}\}$ and a property $Q = \{\text{"very similar ball"}\}$, meaning they are discernible. No matter how similar α and β are, as long as they are different objects, they are discernible by some property. If the two constants denote the same object in the domain, the pair appears in the identity relation I_1^2, and thus the objects are identical. Otherwise, they do not appear in I_1^2 and thus are not identical. Q. E. D. [1]

The Mates rule *I* is not sufficiently powerful to be able to derive this fact in the system, which is why the expansion on rule *I* was added.

Soundness

We will now prove the soundness of the rules of derivation and axioms of $\mathcal{L}_2$. We will begin by simply restating what Benson Mates has presented in his soundness proof, the continuing with the new rules of derivation that we have added. Therefore, our starting point is what we are asserting when we state soundness. We assert the following two things:

1. If ϕ appears on the first line of a derivation, it is the consequence of that line.

2. Any sentence ψ on a later line is a consequence of its premises if all sentences appearing on earlier lines are consequences of theirs.

This leads naturally into a proof by induction. Our base case concerns point 1. There are only 3 cases in which a sentence may appear on the first line:

Case 1, ϕ was entered by rule *P*:

Proof: ϕ has a premise set of $\{1\}$, implying ϕ follows from itself. We know this to be true, therefore the entry is sound.

Case 2, ϕ was entered by rule *T*:

Proof: ϕ has an empty set of premises, implying ϕ is a tautology. Since rule *T* can only enter tautologous lines, the entry is sound.

Case 3, ϕ was entered as an axiom:

Proof: It is necessary to prove soundness for each of the two axioms:

I. In the case of Axiom I (axiom schema of comprehension), the axiom states that there exists a function that is equivalent to any sentence for all variables $\langle x \rangle_n$. This is sound because, under standard semantics, the existential quantifier counts for all properties $P \in \mathcal{P}(d^n)$, in essence, all possible sets of combinations of $\langle x \rangle_n$ are considered. Therefore, any and all possible values of $\phi(\langle x \rangle_n)$ can be expressed by some property.

II. In the case of Axiom II (axiom of choice), the axiom states that if, for any $\langle x \rangle_n$ there exists a y that makes a given property true, there's a function p^n such that for all $\langle x \rangle_n$ the property is true. This is sound because, under standard semantics, the existential quantifier counts for all functions that map tuples in d^n to members of d. Therefore, a function must exist that "selects" the y value necessary to make the given property true, for all $\langle x \rangle_n$.

Our inductive hypothesis will be the following: If a sentence ψ_n is entered on line n, ψ_n is sound. This leads to an inductive step where we try to determine if ψ_{n+1} is sound. Again, we have several cases to consider:

Case 1, ψ_{n+1} was entered by rule *P*:

Proof: ψ_{n+1} has a premise set of $\{n + 1\}$, implying ψ_{n+1} follows from itself. We know this to be true, therefore the entry is sound.

Case 2, ψ_{n+1} was entered by rule *T*:

Proof: If ψ_{n+1} was entered by rule *T*, it adopts the premise lines of all lines cited by rule *T*. Since rule *T* can only enter logical consequences on line $n + 1$ from the premise lines cited, ψ_{n+1} is a logical consequence of all lines less than $n + 1$, therefore the entry is sound.

Case 3, ψ_{n+1} was entered as an axiom:

Proof: This is identical to the proof in the base case, and will not be restated.

Case 4, ψ_{n+1} was entered by rule *C*:

Proof: $\psi_{n+1} = (\phi \rightarrow \chi)$ where χ appears on an earlier line of the derivation. By the inductive hypothesis, χ is a consequence of the premises of that line, which may include ϕ. Therefore, $(\phi \rightarrow \chi)$ (and thus ψ_{n+1}) is a sound entry following from these premises, excluding those of ϕ.

Case 5, ψ_{n+1} was entered by rule *US*:

Proof: $\psi_{n+1} = \phi(\gamma)$ where γ is some first- or second-order constant, and $(\forall\omega)\phi$ appears on some earlier line of the derivation, as cited by *US*. By the inductive hypothesis, $(\forall\omega)\phi$ is a sound entry from its premises. Since $(\forall\omega)\phi \vdash \phi(\gamma)$ by definition of universal generalization, $\phi(\gamma)$ (and therefore ψ_{n+1}) is a consequence of its premises, and therefore a sound entry.

Case 6, ψ_{n+1} was entered by rule *UG*:

Proof: $\psi_{n+1} = (\forall\omega)\phi$ where ω is some first- or second-order variable, and $\phi(\gamma)$ appears on some earlier line of the derivation, as cited by *UG*. By the inductive hypothesis, $\phi(\gamma)$ is a sound entry from its premises. Since $\phi(\gamma) \vdash (\forall\omega)\phi$ if γ appears neither in ϕ nor any premise of ϕ (again, by definition of universal generalization), $(\forall\omega)\phi$ (and therefore ψ_{n+1}) is a consequence of its premises, and therefore a sound entry.

Case 7, ψ_{n+1} was entered by rule *E*:

Proof: Either $\psi_{n+1} = -(\forall\omega) - \phi$ or $\psi_{n+1} = (\exists\omega)\phi$. In the first case, $(\exists\omega)\phi$ appears on an earlier line, and in the second case, $-(\forall\omega) - \phi$ appears on an earlier line. In both cases, by the inductive hypothesis, the sentence appearing on an earlier line is a sound entry, and ψ_{n+1} is a consequence thereof by definition of existential quantification, and therefore a sound entry.

Case 8, ψ_{n+1} was entered by rule *I*:

Proof: If ψ_{n+1} is $\alpha = \alpha$, the entry is sound by definition of the equality relation. If ψ_{n+1} is $\phi(\beta)$ and derives from the lines $\phi(\alpha)$ and $\alpha = \beta$, then by inductive hypothesis, both $\phi(\alpha)$ and $\alpha = \beta$ are sound. Because of the indiscernibility of identicals, $\phi(\alpha)$ is true if and only if $\phi(\beta)$ is true. Therefore, $\phi(\beta)$ is a sound entry. The only other alternative to these two is if ψ_{n+1} is $\alpha = \beta$ and derives from the line $(\forall X)(X\alpha \leftrightarrow X\beta)$. The soundness of this was proved in the previous section. Therefore, ψ_{n+1} is a sound entry.

This proves our system of rules is sound.

Q. E. D. [2]

Incompleteness

The primary drawback of a second-order logic with standard semantics is that it is inherently incomplete. We will establish this fact in this section of the paper making use of Gödel's incompleteness theorems.

Definition 1: Let AR denote the conjunction of the successor, addition, multiplication, and induction axioms. Namely,

$$AR = \big((\forall x)sx \neq 0 \;\&\; (\forall x)(\forall y)(sx = sy \rightarrow x = y)\big) \;\&$$
$$\big((\forall x)x + 0 = x \;\&\; (\forall x)(\forall y)x + sy = s(x + y)\big) \;\&$$
$$\big((\forall x)x \cdot 0 = 0 \;\&\; (\forall x)(\forall y)x \cdot sy = x \cdot y + x\big) \;\&$$
$$(\forall X)\Big(\big(X0 \;\&\; (\forall x)(Xx \rightarrow Xsx)\big) \rightarrow (\forall x)Xx\Big)$$

Definition 2: Let $T = \{\, \phi \mid \phi \text{ is a sentence with no second-order variables and } \vdash AR \rightarrow \phi\}$.

The proof of incompleteness is straightforward and simple: Since the formulas of $\mathcal{L}_2$ are recursively enumerable, so is T. Furthermore, since our deductive system is sound, every sentence of T is true for the natural numbers because each sentence is derivable from AR. From Gödel's incompleteness theorems, the collection of true first-order sentences of arithmetic is not recursively enumerable. So, let ψ be a true sentence of first-order arithmetic not in T. Thus, $\nvdash AR \rightarrow \psi$, but $\vDash AR \rightarrow \psi$ because ψ is a true sentence of first-order arithmetic. Therefore, the sentence $AR \to \psi$ is semantically true, but not provable, meaning our system (and by expansion, any system) of second-order logic under standard semantics is inherently incomplete. Q. E. D. [3]

Since compactness follows from completeness, it is also possible to prove that second-order logic under standard semantics is not compact, i.e. there exists a set Γ that is unsatisfiable, where each finite subset of Γ is satisfiable.

Arbitrary-Order Logic and Sufficiency of Second-Order Logic

We will now introduce the concept of higher-order logics, that is, a language $\mathcal{L}_n$ where $n \geq 2$. This includes second-order logic, but also third-order logic, fourth-order logic, and so on. The first order of business is to discuss our new definition of a well-formed formula. We will first modify our definition of a predicate:

- *Predicates* or *n-order predicate constants* are represented by the upper-case letters 'A' through 'T,' with a subscript denoting its order, possibly with a numerical superscript denoting its degree. That is, P_2^3 is a second-order predicate of degree 3, and B_5 represents a fifth-order predicate. Note that in this expansion, subscripts have significant meaning, and are not compatible with the language $\mathcal{L}_2$. An alternative is to use a left-handed superscript to denote order (so, for example, ${}^5B$ instead of B_5).
- *Predicate variables* are represented by the upper-case letters 'U' through 'Z,' with a subscript denoting its order and possibly with a numerical superscript denoting its degree.

Furthermore, we modify our definition of an atomic sentence:

- *Atomic formulas* are crafted in any of the following ways:
    a. A sentential letter is an atomic formula.
    b. If α and β are individual symbols, $\alpha = \beta$ and $\alpha \neq \beta$ are atomic formulas.

c. A predicate of degree n and order 2 followed by n individual symbols is an atomic formula.

d. A predicate of degree n and order $m + 1$ followed by n m-order predicates is an atomic formula. For example, $P_4^2 A_3 B_3$ is an atomic formula.

This leaves us able to craft sentences such as $(\forall X_3)X_3P_2$, meaning the predicate P_2 exists in every third-order predicate.

The next step is to discuss semantics. Assuming standard semantics and a model $M = \langle d, I\rangle$, then we know that each second-order n-degree predicate ranges over $\mathcal{P}(d^n)$. Each third-order n-degree predicate will range over $\mathcal{P}\left(\left(\bigcup_{i=1}^{\infty}\mathcal{P}\left(d^i\right)\right)^n\right)$, that is, the powerset of the nth cross product of the union of all second-order predicates. Each fourth-order n-degree predicate ranges over the powerset of the nth cross product of the union of all third-order predicates, represented by $\mathcal{P}\left(\left(\bigcup_{j=1}^{\infty}\mathcal{P}\left(\left(\bigcup_{i=1}^{\infty}\mathcal{P}\left(d^i\right)\right)^j\right)\right)^n\right)$ and so on. This pattern extends upwards to n-order, m-degree predicates.

This makes higher-order logic extremely expressive. For example, one can state the metaphysics principle of "if two items are alike in all physical respects, they are alike in all mental respects" very clearly in third-order logic, using the following construction: Let P_3 be the set of all physical properties and M_3 be the set of all mental properties. The concept is expressed as:

$$(\forall X)(\forall Y)(\forall x)(\forall y)\Big((P_3X \,\&\, M_3Y) \rightarrow \big((Xx \leftrightarrow Xy) \rightarrow (Yx \leftrightarrow Yy)\big)\Big)$$

A direct translation is, "for all properties X and Y, and for all pairs of items x and y, if X is a physical property and Y is a mental property, then the following is true: if x and y are alike for all properties X, then x and y are alike for all properties Y," or in a more down-to-earth manner, "if two objects are alike in all physical respects, they are alike in all mental respects."

By the way I have been talking about higher-order logic, one would think as you go up in order, you would gain more expressive power. However, this is not the case. The concept of *reducibility to second-order logic* states that any n-order sentence of logic, where $n \geq 3$, can be reduced, or rewritten, to an equivalent second-order sentence of logic. This is precisely what makes second-order logic so special and of so much interest. It is sufficient to use second-order logic and not need to rely upon higher-order logic, but as we will see, second-order logic is irreducible to first-order logic.

Second-Order Logic is Necessary

As alluded to, while all higher-order logic is reducible to second-order logic, the same is not true down to first-order logic. This means that quantification over predicates adds strictly more expressiveness.

Proof: suppose there existed a procedure by which we can reduce any given second-order sentence order to the first order. Let S be a set of unsatisfiable sentences in second-order logic, of which every finite subset is satisfiable. This is possible because compactness is not preserved. By reducing every sentence to the first-order, we are left with a new set of sentences S'. By compactness of first-order logic, either some finite subset of S' is unsatisfiable, or S' is satisfiable. In either case, the semantics of the sentences have changed and the reduction failed. Therefore, it is impossible to reduce second-order logic to the first order. Q. E. D. [4]

Now that we have proven that second-order logic provides more expressiveness, it is time to weigh the benefits vs. the cost. The biggest problem people tend to have with second-order logic is its loss of completeness and associated properties, such as compactness. This is indeed a very large loss, and not one to scoff at. The benefits, on the other hand, are also quite clear: the ability to quantify over predicates gives substantial expressive power to the logic. The ability to not only talk about arbitrary properties, but also specific types of properties, is incredibly helpful in many instances. In many cases, proofs in higher-order logic are far more elegant than the alternatives.

In the end, use the best tool for the job. One should try to restrict themselves to the lowest-order logic needed to solve a given problem, starting with the propositional calculus, then moving up to first-order predicate calculus, and finally up to second-order predicate calculus, higher-order predicate calculus, or alternatives. This ensures maximal completeness and decidability for the problem at hand, as first-order logic loses decidability and second-order logic loses completeness.

Second-order logic is sometimes questioned at its foundations, though, in an argument referred to as *foundationalism*, hence the title of Shapiro's book. Philosophers like Willard Van Orman Quine deny that second-order logic is a logic in the first place, and is rather a set theory in disguise. The foundational argument made is that a substantial amount of mathematics can be expressed in second-order logic, so attempting to draw a border between logic and mathematics would separate second-order logic into the realm of mathematics, not logic. Shapiro argues that such a distinction need not be made. As for the claim that the loss of completeness, compactness, Löwenheim–Skolem, etc. are "defects" in second-order logic, Shapiro further argues that these are natural sacrifices to the inherent strength of the logic.

Conclusion

The takeaway to be made is that our particular expansion of $\mathcal{L}'$ to $\mathcal{L}_2$ is a viable expansion of first-order logic to second-order logic, and is sufficiently powerful for a new set of problems.

Standard semantics were elaborated on, and Henkin semantics were introduced, and many properties were then established about our system, including identity of indiscernibles, soundness, completeness, upward reducibility, and downward irreducibility. The idea of second-order logic should be in every logician's standard toolbox, and it is dismissive to ignore the power of such a system.

Notes

[1] – This proof of identity of indiscernibles under standard semantics is entirely my own proof. I am not saying the proof doesn't exist elsewhere, but I independently discovered and wrote it up.

[2] – The soundness proof was largely taken from Mates, modified only to add the axioms and rules of the new system.

[3] – The incompleteness proof was taken from Shapiro in *Foundations without Foundationalism*, explained in my own words.

[4] – This irreducibility proof is entirely my own proof. I am not saying the proof doesn't exist elsewhere, but I independently discovered and wrote it up.

References

Mates, B. (1972). *Mates Elementary Logic.* Oxford University Press.

Shapiro, S. (2002). *Foundations without Foundationalism.* Oxford University Press.