On Mathematical Induction

Nikolai I. Kazimirov

2026

Contents

1 Formalizations and Varieties of Induction

The principle of mathematical induction is a load-bearing construction of regular discrete systems. Depending on the expressive power of the language and the deductive strength of the base theory, induction takes on substantially different logical forms.

1.1 First-Order and Second-Order Formulations

In first-order Peano arithmetic (PA), induction cannot be expressed by a single formula, because quantification over predicates is unavailable: formally, it is postulated as a countable scheme of axioms, one for each formula $\varphi(a)$ (possibly with parameters $\vec{p}$):

$$\varphi[a/0] \wedge \forall x(\varphi[a/x] \to \varphi[a/Sx]) \to \forall y\ \varphi[a/y].$$

In substance, however, it is more familiar and convenient to write this scheme in the form mathematicians are used to from inductive definitions in set theory — identifying the formula φ with its extension $X = \{x \mid \varphi(x)\}$:

$$I: \qquad \forall X \Big(0 \in X \ \wedge \ \forall x \, (x \in X \to Sx \in X) \ \to \ \forall x \, (x \in X) \Big). \tag{1}$$

One must immediately clarify the sense in which the notation $\forall X$ in (1) is to be understood: in the first-order language of PA there is no quantification over sets, so (1) is a *metanotation* for the axiom scheme (ranging over all formulas φ defining X), and not a single formula with a genuine second-order quantifier. It is exactly this scheme — denoted I, in agreement with the already-standard notation $I\Sigma_n$, $I\Delta_0$ for its fragments by the complexity of the formula φ — that we discuss in detail in §2, together with several principles equivalent to it.

In the system of second-order arithmetic (Z_2), where quantification over sets is permitted directly by the language, scheme (1) collapses into a single literal axiom of the same form — now without any qualification about metanotation:

$$\forall X \, (0 \in X \wedge \forall x (x \in X \to S(x) \in X) \to \forall y (y \in X)) \, . \tag{2}$$

The deductive strength of Z_2 is determined by which *comprehension scheme* is postulated for the existence of sets $X = \{x \mid \varphi(x)\}$: the wider the class of formulas φ for which the existence of such an X is guaranteed, the larger the stock of "sets" to which the single axiom (2) applies. Under impredicative comprehension (adopted in full Z_2), this already covers an arbitrary subset definable in the whole language of analysis.

1.2 Varieties of Induction by Formula Complexity and Parameters

1. **Restriction by formula hierarchy ($I\Sigma_n$, $I\Pi_n$):** Restricting the quantifier complexity of the formula φ in the induction scheme generates a strict deductive hierarchy of arithmetic subsystems. For instance, $I\Sigma_1$ suffices to prove the totality of the primitive recursive functions, whereas $I\Delta_0$ (induction only for formulas with bounded quantifiers) cannot even prove the totality of exponentiation.

2. **Induction with and without parameters (parameter-free induction):** In full PA, the presence of free variables $\vec{p}$ in the formula φ does not increase the strength of the theory. However, for restricted fragments such as $I\Sigma_n^-$ (without parameters), the strength of the theory is strictly less than for the fragments with parameters $I\Sigma_n$.

3. **Induction along orderings:** A generalization of the local step to arbitrary well-ordered sets.

1.3 Induction in Set Theory

In Zermelo–Fraenkel set theory (ZFC), induction on the natural numbers is a consequence of the axiom of infinity, which guarantees the existence of an inductive set ω:

$$\exists X \, (\emptyset \in X \wedge \forall x \in X (x \cup \{x\} \in X)) \, . \tag{3}$$

The set ω is defined as the intersection of all such inductive sets, from which standard induction for subsets of ω follows trivially.

Transfinite induction generalizes this principle to the class of all ordinals On (or to arbitrary well-founded relations):

$$\forall \alpha \in \mathrm{On}\,((\forall \beta < \alpha\ \varphi(\beta)) \to \varphi(\alpha)) \to \forall \gamma \in \mathrm{On}\ \varphi(\gamma). \tag{4}$$

In ZFC, transfinite induction as a proof principle (as opposed to transfinite *recursion*, for defining functions by which the replacement scheme is essential) is a theorem, resting only on the axiom of regularity (foundation) and the separation scheme.

2 Equivalent First-Order Principles and Metamathematical Subtleties

2.1 Five Principles: Definitions

Consider five key principles, formulated in the first-order language of arithmetic. As in §1.1, each of them is an axiom scheme (over the formula φ defining $X = \{x \mid \varphi(x)\}$), and the notation $\forall X(\ldots)$ everywhere denotes exactly this scheme, not a second-order quantifier. For x, write $[0, x) \overset{\text{def}}{\rightleftharpoons} \{y \mid y < x\}$ for the initial interval, and fix three abbreviations:

$$\mathrm{Ind}(X) \overset{\text{def}}{\rightleftharpoons}\ 0 \in X\ \wedge\ \forall x\,(x \in X \to Sx \in X),$$

$$\mathrm{Prog}(X) \overset{\text{def}}{\rightleftharpoons}\ \forall x\,\big([0, x) \subseteq X\ \to\ x \in X\big),$$

$$\mathrm{Tot}(X) \overset{\text{def}}{\rightleftharpoons}\ \forall x\,(x \in X).$$

Here *inductiveness* is the local-step condition; *progressiveness* is closure with respect to initial intervals; *totality* states that X contains absolutely everything. Then:

1. **Local induction (I).**
   $\forall X\big(\mathrm{Ind}(X) \to \mathrm{Tot}(X)\big)$.

2. **Course-of-values (complete) induction (I^*).**
   $\forall X\big(\mathrm{Prog}(X) \to \mathrm{Tot}(X)\big)$.

3. **Least element principle (L).**
   $\forall X\big(X \neq \varnothing\ \to\ \exists x \in X\,\forall y \in X\,(x \leqslant y)\big)$.

4. **Infinite descent principle (D).**
   $\forall X\big(X \neq \varnothing\ \to\ \neg\big(\forall x \in X\,\exists y \in X\,(y < x)\big)\big)$.

5. **Collection scheme (B).**
   $\forall a\,\big(\forall x{<}a\,\exists y\,\varphi(x, y)\ \to\ \exists z\,\forall x{<}a\,\exists y{<}z\,\varphi(x, y)\big)$.

(It is inconvenient and unnatural to write the scheme B in terms of a characteristic set X — by its very form it is a statement about a binary relation $\varphi(x, y)$ — so for it we keep the ordinary formula notation.)

2.2 Equivalence of I^*, L, D in Pure Predicate Calculus

A principally important observation: I^*, L, and D are equivalent to one another already in pure predicate calculus — without a single arithmetical axiom — provided only that $<$ is required to be a strict linear order. We do not need a detailed axiomatics of order here (that is not our goal); only one consequence of it matters — *trichotomy*: for any x, y, exactly one of $x < y$, $x = y$, $y < x$ holds. Trichotomy is used exactly once: to link the "negative" form of the principle D (expressed via $<$) with its logical twin — the "positive" form L familiar to a mathematician (via $\leqslant$, comparison with *every* element of X).

Proposition 1. *For any formula φ (with $X = \{x \mid \varphi(x)\}$):*

(a) in pure predicate calculus, $I^*_{\neg\varphi} \iff D_\varphi$;

(b) if, in addition, $<$ satisfies trichotomy, then $D_\varphi \iff L_\varphi$.

Since the class of formulas is closed under $\varphi \mapsto \neg\varphi$, it follows that the full schemes are equivalent: $I^* \iff D \iff L$.

Proof. (a) The scheme I^* for the formula $\neg\varphi$ has the form $A \to B$, where (unfolding Prog and Tot by definition)

$$A \equiv \forall x \big(\forall y{<}x\, (\neg\varphi(y)) \to \neg\varphi(x)\big), \qquad B \equiv \forall x\, (\neg\varphi(x)).$$

By the law of contraposition, $A \to B$ is equivalent to $\neg B \to \neg A$ — a pure propositional tautology, requiring no new axioms whatsoever. Unfolding the negations by De Morgan's rules and quantifier duality:

$$\neg B \equiv \exists x\, (\varphi(x)) \equiv X \neq \varnothing, \qquad \neg A \equiv \exists x \big(\varphi(x) \wedge \forall y{<}x\, (\neg\varphi(y))\big).$$

Inside $\neg A$, the subformula "$\forall y{<}x\, \neg\varphi(y)$", that is, "$\forall y\, (y < x \to \neg\varphi(y))$", is itself, by pure logic, equivalent to its contrapositive "$\forall y\, (\varphi(y) \to \neg(y < x))$", that is, "$\forall y \in X\, \neg(y < x)$". Hence

$$\neg A \equiv \exists x \in X\, \forall y \in X\, \neg(y < x) \equiv \neg\big(\forall x \in X\, \exists y \in X\, (y < x)\big),$$

and the implication $\neg B \to \neg A$ is exactly D_φ.

(b) Unfolding the inner negation, D_φ states: $X \neq \varnothing \to \exists x \in X\, \forall y \in X\, \neg(y < x)$. The formula L_φ is literally the same, only with "$\neg(y < x)$" replaced by "$x \leqslant y$". By trichotomy these are the same condition: $\neg(y < x)$ means that either $y = x$ or $x < y$, i.e., exactly $x \leqslant y$. □

Thus the equivalence $I^* \iff L \iff D$ is a fact of *logic*, not of arithmetic: the proof uses neither the number 0 nor the operation $+1$, only the linearity of the order $<$.

2.3 Equivalence of I and I^*: the Role of Minimal Arithmetic of the Successor

The connection of local induction I with the triple I^*, L, D is of a fundamentally different nature: it is inseparable from the successor S and therefore cannot be a purely logical fact. We will need three elementary facts about $<$ and S. Here one should be careful about where these come from: in Robinson arithmetic Q, the relation $<$ is usually not primitive at all, but is *defined* by an existential formula ($x < y \stackrel{\text{def}}{\rightleftharpoons} \exists z\, (Sz + x = y)$), and the proof that this

definition is equivalent to the recursive identities given below is itself not entirely trivial and depends on the manner of recursively unfolding $+$. To avoid leaving this hanging, we fix $<$ as a primitive relation and introduce the theory

$$\mathsf{PA}^{\mathrm{rec}}_{-} \stackrel{\text{def}}{\rightleftharpoons} \big\{\, x+0=x,\ x+Sy=S(x+y),\ x\cdot 0=0,\ x\cdot Sy=x\cdot y+x,$$

$$Sx\neq 0,\ Sx=Sy\to x=y \,\big\} \cup \{\text{(O1)–(O3)}\}$$

— essentially, this is Robinson arithmetic in its original, recursive form, with the order axioms (O1)–(O3) explicitly postulated (rather than derived):

(O1) $\neg(x<0)$ — 0 is the least element;

(O2) $y<Sx \leftrightarrow y\leqslant x$ — the recursive link between order and successor;

(O3) $x\neq 0\to\exists y\,(x=Sy)$ — every nonzero element is a successor.

All three are axioms of $\mathsf{PA}^{\mathrm{rec}}_{-}$, not theorems derivable from something simpler.

Lemma 2. *From (O1)–(O3) it follows that: for every X,* $\mathrm{Ind}(X)\to\mathrm{Prog}(X)$.

Proof. Suppose $\mathrm{Ind}(X)$ holds, i.e., $0\in X$ and $\forall z\,(z\in X\to Sz\in X)$. Take an arbitrary x with $[0,x)\subseteq X$; we show $x\in X$. By (O3), either $x=0$ or $x=Sz$ for some z. If $x=0$: then $x\in X$ directly, by the first conjunct of $\mathrm{Ind}(X)$. If $x=Sz$: then by (O2) (with $y=z$) we have $z<Sz=x$, whence $z\in[0,x)\subseteq X$, i.e., $z\in X$; then by the second conjunct of $\mathrm{Ind}(X)$ we get $Sz=x\in X$. □

Proposition 3. *Given (O1)–(O3), I^* implies I.*

Proof. Suppose $\mathrm{Ind}(X)$ holds. By the Lemma, $\mathrm{Prog}(X)$. Applying I^*, we obtain $\mathrm{Tot}(X)$ — the conclusion of I. □

Note: axiom (O3) is used here exactly once — in the case split on x (either 0 or a successor). Without it, one could not guarantee that every nonzero element is "reachable by an induction step" from some predecessor.

Proposition 4. *Given only (O1)–(O2) (axiom (O3) is not needed!), I implies I^*.*

Proof. Suppose $\mathrm{Prog}(X)$ holds. Define the auxiliary set

$$Y \stackrel{\text{def}}{\rightleftharpoons} \{x \mid [0,x)\subseteq X\}$$

— the maximal initial interval lying entirely within X.

$Y\subseteq X$. If $x\in Y$, then $[0,x)\subseteq X$ — which is exactly the hypothesis of $\mathrm{Prog}(X)$ at this x, so immediately $x\in X$.

$\mathrm{Ind}(Y)$. *Base:* by (O1), $[0,0)=\varnothing\subseteq X$, so $0\in Y$. *Step:* let $z\in Y$, i.e., $[0,z)\subseteq X$; by the already-proved $Y\subseteq X$, also $z\in X$, so $[0,z)\cup\{z\}\subseteq X$. By (O2), $[0,Sz)=[0,z)\cup\{z\}$ (since $y<Sz\leftrightarrow y\leqslant z\leftrightarrow y<z\ \vee\ y=z$), so $[0,Sz)\subseteq X$, i.e., $Sz\in Y$.

Applying I to Y, we obtain $\mathrm{Tot}(Y)$. Together with the already-established $Y\subseteq X$, this immediately gives $\mathrm{Tot}(X)$ — the conclusion of I^*. □

The asymmetry of the directions here is substantive. The direction $I^* \Rightarrow I$ (the Lemma above) essentially uses all three facts (O1)–(O3), with (O3) playing the decisive role — in the case split "x is 0 or a successor". The direction $I \Rightarrow I^*$, by contrast, gets by with (O1)–(O2), and (O3) is simply not used in its proof.

(O3) is essential: a model where I^* holds but I does not. Take as carrier the ordinal $\omega + \omega$ — two consecutive copies of $\mathbb{N}$, with the usual ordinal order $<$, and define the successor S *separately within each copy*: $S(n) = n+1$ in the first copy, $S(\omega+n) = \omega+n+1$ in the second. The function S is everywhere defined and injective, (O1)–(O2) hold, but the element ω (the start of the second copy) is not $S(y)$ for any y — this is exactly the limit ordinal at which (O3) fails.

Since $\omega+\omega$ is a genuine ordinal, i.e., a well-ordered set, transfinite induction (§1.3) holds in it for *arbitrary* subsets of the carrier — not only for subsets given by a formula — so I^* (and with it L and D) holds here in a genuinely second-order sense, with no reservations about a scheme.[1] But I does not hold: the set X = the first copy of $\mathbb{N}$ contains 0 and is closed under S (i.e., $\mathrm{Ind}(X)$ holds), yet $X \neq \omega + \omega$. This is exactly what the asymmetry above predicts: the model violates (O3) — and with it I — but does not touch the purely logical triple I^*, L, D at all.

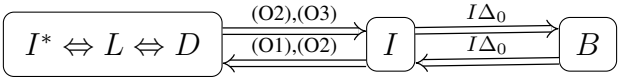
2.4 The Collection Scheme B and Induction

The nature of the connection between the collection scheme B and the induction scheme I is different, and noticeably more subtle metamathematically, than everything considered above: it reduces neither to pure logic nor to the minimal arithmetic of the successor. We will not prove it here, confining ourselves to a precise statement of the known result.

Within the weak base theory PA^- (the axioms of a discrete ordered semiring without induction, see Kaye, *Models of Peano Arithmetic*), the scheme B on its own **does not imply** the scheme I. But already relative to the base induction for Δ_0-formulas ($I\Delta_0$), the connection is restored and takes the form of a strict hierarchy (Paris–Kirby, 1978):

$$I\Sigma_n \;\to\; B\Sigma_n \;\to\; I\Sigma_{n-1} \qquad (n \geq 1),$$

while at the level of the full (formula-complexity-unbounded) schemes, $I \iff B$ relative to the base $I\Delta_0$. The hierarchy of connections among all five principles is shown in the diagram:



The equivalence $I^* \Leftrightarrow L \Leftrightarrow D$ is proved in §2.2 purely logically, up to the trichotomy of $<$. The passage to I requires the minimal arithmetic of the successor (§2.3): it is precisely axiom (O3) that creates the asymmetry — it is needed for $I^* \Rightarrow I$, but not for $I \Rightarrow I^*$. The passage to the collection scheme B essentially uses the induction $I\Delta_0$; the proof is not given.

[1] A remark for the cautious reader: does this not mean that induction "forces" a model to be well-ordered, i.e., standard? No: there exist nonstandard models of PA in which the full scheme I (and with it I^*, L, D as schemes) holds, yet the model itself, viewed *from outside*, is not well-ordered — it contains subsets (not definable in the language of PA) with no least element. There is no contradiction here: as an axiom scheme, I^* guarantees a least element only for subsets *definable* by a formula of the language of PA; the scheme simply does not extend to the "illegitimate" subsets of a nonstandard model. In the example above we were legitimately entitled to require I^* for absolutely all subsets of the carrier precisely because $\omega + \omega$ is standard by construction (coincides with a genuine ordinal); for nonstandard models this luxury is no longer available.

3 Consequences of the Induction Principle in Mathematical Structures

Below is a list of deep mathematical statements whose validity is essentially tied to the deductive strength of induction. Where possible, the exact status of the statement is indicated in terms of subsystems of second-order arithmetic (RCA_0, WKL_0, ACA_0).

3.1 The Pigeonhole Principle

Functional formulation: if A and B are finite sets, and the cardinality of A is strictly greater than the cardinality of B, then there is no injective map $f : A \to B$. *Status:* The provability of the pigeonhole principle is a classical question of proof theory. By the theorem of Paris–Wilkie–Woods and Ajtai (1988), PHP is not provable in $I\Delta_0$ even for Δ_0-definable functions — that is, one of the most elementary combinatorial truths is not derivable from bounded induction, which demonstrates the extreme weakness of $I\Delta_0$ as a base theory.

3.2 König's Lemma on Infinite Paths

Formulation in the language of partially ordered sets: if a finitely branching partially ordered set $(T, \leqslant_T)$ that is a tree with a single root contains infinitely many elements, then it contains an infinite chain (path). *Status:* Weak König's lemma (WKL_0, for binary trees) is equivalent to the compactness of an interval in analysis. Full König's lemma (for arbitrary finitely branching trees) is strictly stronger and equivalent to the arithmetical comprehension scheme ACA_0. The extra strength compared to WKL_0 arises not from induction (the induction scheme is the same in both systems), but precisely from the comprehension scheme: to choose, at each step, a subtree with infinitely many vertices, one must be allowed the predicate "the subtree is infinite," which for arbitrary (not just binary) trees does not reduce to a Δ_0-condition and requires exactly ACA_0.

3.3 Algebraic Laws in PA

Commutativity and associativity of addition ($\forall x, y\ x + y = y + x$) and of multiplication.

Clarifying the base theory. It is important here to specify precisely which induction-free theory is meant. The theory PA^-, used in §2.4 (the theory of the nonnegative part of a discrete ordered ring, see Kaye, *Models of Peano Arithmetic*), already *postulates* the commutativity and associativity of $+$ and $\times$ as separate, non-inductive axioms — so one cannot speak of "noncommutative models of PA^-": that would contradict the very definition of PA^-. The statement below concerns the weaker theory $\mathsf{PA}^{\mathrm{rec}}_-$, introduced in §2.3 — Robinson arithmetic in its original, recursive form, with explicit order axioms (O1)–(O3), but without induction and without separately postulating commutativity or associativity.

A model without commutativity: Deriving the commutative law in $\mathsf{PA}^{\mathrm{rec}}_-$ without induction is impossible: there exist (necessarily nonstandard) models of $\mathsf{PA}^{\mathrm{rec}}_-$ in which addition is not commutative. A substantive illustration of the same phenomenon — though not a literal model of $\mathsf{PA}^{\mathrm{rec}}_-$, but a structure of a different type (an ordered monoid) — is ordinal arithmetic: for the first transfinite ordinal ω, $\omega + 1 \neq 1 + \omega$ holds. It is precisely the impossibility of deriving commutativity without induction that is the reason why PA^- in the sense of Kaye includes commutativity and associativity explicitly in its list of axioms, rather than trying to obtain them as consequences.

3.4 Dedekind-Finite Sets

1. **Proper subsets:** The statement "if $B \subsetneq A$ and A is finite, then the cardinality of B is strictly less than the cardinality of A" is a direct consequence of induction on the number of elements of the set A. Without induction, finite sets behave like quasi-infinite ones.

2. **Dedekind finiteness:** A set A is Dedekind-finite if there is no injection $f : A \to A$ that fails to be a surjection. Using induction, one easily proves the direction "standard-finite $\Rightarrow$ Dedekind-finite." The converse direction, however, does not follow from induction alone: it is provable using the countable axiom of choice AC_ω (if the set is infinite, AC_ω allows choosing a countable subsequence of distinct elements within it, which yields the required non-surjective injection) — but this is only a *consequence* of AC_ω, not a statement equivalent to it. In any case, without any choice, in the theory ZF this direction is unprovable: there exist models of ZF with infinite Dedekind-finite sets (Fraenkel–Mostowski models, Cohen's first model).

3.5 Ramsey's Theorem (RT^2_2)

Ramsey's theorem for a pair of colors is most clearly stated in the language of ordinary graphs, without talk of colorings: for two colors, one of them is "there is an edge" and the other is "there is no edge," so an arbitrary coloring of pairs of vertices in 2 colors is the same thing as an arbitrary graph on those vertices.

The finite case. The classical example is the "party theorem": among any 6 people (consider the acquaintance graph, where an edge is an acquaintance), there are either three who are pairwise acquainted (a triangle in the graph) or three who are pairwise unacquainted (a triangle in its complement); 6 is the smallest such number (the Ramsey number $R(3,3) = 6$). In general: for any k, m there exists a Ramsey number $R(k,m)$ — the smallest N such that any graph on N vertices contains either a clique of size k or an independent set of size m. It is worth noting that even this purely finitary statement (merely quantified over k, m) is proved by induction, on the sum $k + m$.

The infinite case (RT^2_2). In any graph on an infinite vertex set there is an infinite subset of vertices forming either a complete subgraph (all edges present) or a totally disconnected subgraph (no edges at all). This is a direct consequence of induction (more precisely, of the principle L: among the "types" of vertices relative to those already selected, there are always infinitely many of some single type).[2]

3.6 Totality of the Ackermann Function

By Parsons's theorem, the $I\Sigma_1$-provably recursive functions are exactly the primitive recursive functions; since the Ackermann function is by construction not primitive recursive, its totality is **unprovable** in $I\Sigma_1$. It becomes provable already at the next level, in $I\Sigma_2$ (which corresponds to the place of the Ackermann function in the fast-growing hierarchy, at level ω, whereas $I\Sigma_1$ covers only the finite levels of this hierarchy).

[2]The exact reverse-mathematical strength of RT^2_2 is known and quite subtle: it is provable in $\mathsf{RCA}_0 + I\Sigma_2$, but does not itself imply $I\Sigma_2$ (Cholak–Jockusch–Slaman, 2001) — we will not go into these details.

3.7 The Limits of Induction: Goodstein's Theorem and the Paris–Harrington Principle

All the examples of §§3.1–3.6 showed how much can be achieved by the means of induction I. The next two results demonstrate the exact opposite: induction I (that is, the full induction scheme of PA) has a clear, precisely measurable **limit** of strength — and it is exactly these two examples that mark it.

Goodstein's theorem. Formulated and proved (with no connection whatsoever to the metamathematics of PA) by R. Goodstein as early as 1944: for any number n, the Goodstein sequence $g_n(0), g_n(1), g_n(2), \ldots$ (defined via the representation of numbers in hereditary base-k notation, replacing the base k by $k+1$ and subtracting one at each step) always reaches 0 in finitely many steps. This is a true Π_2-sentence of the language of PA (of the form $\forall n\, \exists k\, g_n(k) = 0$); it was only almost four decades later that Kirby and Paris (1982) showed it is **not derivable in PA**.

The Paris–Harrington principle. A strengthened version of the finitary Ramsey theorem of §3.5 (with the additional requirement of "relative largeness" of the sought monochromatic set) — a true, purely combinatorial statement (Paris, Harrington, 1977), which was presented *immediately*, at the very moment of its appearance, as an example of a statement unprovable in PA. In this sense the Paris–Harrington principle is the first example of a *natural* theorem (not metamathematical in its statement, not explicitly appealing to provability or to Gödelian diagonalization) for which unprovability in PA was established simultaneously with the theorem itself: Goodstein's theorem as a *mathematical statement* appeared considerably earlier (1944), but its independence from PA was discovered only later — under the influence of the Paris–Harrington result.

Both facts are nonetheless provable by means going beyond I — in particular, already in ZFC and in full second-order arithmetic Z_2: it suffices to use *transfinite* induction along the ordinal ε_0, available already in the comparatively modest subsystem $\mathsf{ATR}_0 \subseteq \mathsf{Z}_2$. The ordinal ε_0 does not appear here by accident: by Gentzen's theorem (1936), this is exactly the proof-theoretic ordinal of PA — the boundary that induction I cannot formalize as a proof scheme on its own. In other words: the local and course-of-values induction schemes $I \iff I^*$ — for all the power they demonstrate — are themselves **not closed** with respect to all true arithmetical statements; going beyond their limits requires an essentially stronger form of induction — transfinite induction — unavailable as a first-order axiom scheme over the standard model ω.

List of Notation

PA	Peano arithmetic (first-order, with the full induction scheme I).
Z_2	Second-order arithmetic (analysis) with the impredicative comprehension scheme.
ZFC	Zermelo–Fraenkel set theory with the axiom of choice.
PA^-	The axioms of a discrete ordered semiring without induction (in the sense of Kaye), with commutativity and associativity of $+, \times$ explicitly postulated.

$\mathsf{PA}^{\mathrm{rec}}_{-}$	Robinson arithmetic in recursive form (equations for $+, \times, S$) with explicit order axioms (O1)–(O3), without induction and without separately postulating commutativity/associativity (§2.3).
$S, <, \leqslant$	Successor; strict and non-strict order on numbers.
$[0, x)$	The initial interval $\{y \mid y < x\}$.
$\mathrm{Ind}(X)$	Inductiveness of X: $0 \in X \ \wedge \ \forall x\,(x \in X \to Sx \in X)$.
$\mathrm{Prog}(X)$	Progressiveness of X: $\forall x\,([0, x) \subseteq X \to x \in X)$.
$\mathrm{Tot}(X)$	Totality of X: $\forall x\,(x \in X)$.
I	Local induction: $\forall X(\mathrm{Ind}(X) \to \mathrm{Tot}(X))$.
I^*	Course-of-values (complete) induction: $\forall X(\mathrm{Prog}(X) \to \mathrm{Tot}(X))$.
L	The least element principle.
D	The infinite descent principle (absence of infinite descending chains).
B	The collection scheme.
(O1)–(O3)	Axioms of $\mathsf{PA}^{\mathrm{rec}}_{-}$ relating $<$ and S: 0 is least; $y < Sx \leftrightarrow y \leqslant x$; every nonzero element is a successor.
$I\Sigma_n$, $B\Sigma_n$, $I\Delta_0$	Induction (resp. collection) for formulas of bounded quantifier complexity; the hierarchy of subsystems of PA.
RCA_0	The base system of reverse mathematics: recursive comprehension scheme plus $I\Sigma_1$.
WKL_0	RCA_0 plus weak König's lemma (for binary trees).
ACA_0	RCA_0 plus the arithmetical comprehension scheme.
ATR_0	RCA_0 plus arithmetical transfinite recursion; strictly stronger than ACA_0.
ε_0	The proof-theoretic ordinal of PA (Gentzen, 1936): the supremum of the order types of canonical (naturally given) well-ordering relations on $\mathbb{N}$ whose well-foundedness is provable in PA.[3]
RT^2_2	Ramsey's theorem for pairs and two colors (in graph form — for an arbitrary graph on an infinite vertex set).

[3] Not the least ordinal for which PA cannot prove well-foundedness: there exist far shorter orders, artificially built (via the consistency predicate) with the same defect. The qualification "canonical" is precisely meant to exclude such pathologies — the question of the precise boundary of this notion is the subject of a separate discussion in proof theory.