Frege: Logical objects by abstraction and their criteria of identity

Matthias Schirn (University of Munich, Munich Center of Mathematical Philosophy)

Abstraction à la Frege

A schema for a Fregean abstraction principle can be stated as follows:

(S) $Q(\alpha) = Q(\beta) \leftrightarrow R_{eq}(\alpha, \beta)$.

Here "Q" is a singular term-forming operator, α and β are free variables of the appropriate type, ranging over the members of a given domain, and "R_{eq}" is the sign for an equivalence relation holding between the values of α and β. In *The Foundations,* the paradigms of Fregean abstraction are:

(1) $D(a) = D(b) : \leftrightarrow a \parallel b$.

The direction of line *a* is identical with the direction of line *b* if and only if line *a* is parallel to line *b*.

(2) $N_x F(x) = N_x G(x) \leftrightarrow Eq_x(F(x),G(x))$.

The number of *F*s is identical with the number *G*s if and only if *F* and *G* are equinumerous, i.e., according to Frege's definition of equinumerosity in §72 of *The Foundations*: if and only if there is a relation *R* that correlates one-to-one the *F*s and the *G*s. As usual, I refer to (2) as "Hume's Principle".

In *Basic Laws*, Frege's paradigm is

(3) $\dot{\varepsilon}f(\varepsilon) = \dot{\varepsilon}g(\varepsilon) \leftrightarrow \forall x(f(x) = g(x))$.

The course-of-values of the function *f* is identical with the course-of-values of the function *g* if and only if *f* and *g* are coextensive.

Principle (3) is the famous-infamous Axiom V of Frege's logical system in *Basic Laws*, the exact structural analogue of (2) .

As to Frege's idea towards the end of *The Foundations* that extensions of concepts could be dispensed with in pursuit of the logicist programme, there are just two options that he might have had in mind: (a) to identify the cardinals with objects other than extensions of concepts

or (b) to resume the tentative contextual definition of the cardinality operator (see below) and make an additional stipulation without invoking extensions of concepts. I analyze (a) and (b).

Hume's Principle and higher-order abstraction in The Foundations

In order to prove Hume's Principle, Frege has to show, according to the final explicit definition of the cardinality operator (see below), that (1) $Eq_x(F(x),G(x)) \rightarrow \#\xi(Eq_x(\xi(x),F(x))) = \#\xi(Eq_x(\xi(x),G(x)))$; that is to say, he has to prove that "under this hypothesis" the following two sentences hold generally: (2) $Eq_x(H(x),F(x)) \rightarrow Eq_x(H(x),G(x))$ and (3) $Eq_x(H(x),G(x)) \rightarrow Eq_x(H(x),F(x))$. Thus, Frege is converting here the statement that the extensions of two specific second-level concepts are identical into the statement that these concepts are coextensive: $\#\xi(Eq_x(\xi(x),F(x))) = \#\xi(Eq_x(\xi(x),G(x))) \leftrightarrow \forall\xi(Eq_x(\xi(x),F(x)) \leftrightarrow Eq_x(\xi(x),G(x)))$. We might thus presume that in introducing extensions of second-level concepts he had in mind the following third-order abstraction principle (modelled on Axiom V):

$$\#f(\Phi\Psi(f(\Phi\Psi))) = \#f(\Phi\Psi(f(\Phi\Psi))) \leftrightarrow \forall f(\Phi\Psi(f(\Phi\Psi)) \leftrightarrow \Phi\Psi(f(\Phi\Psi))).$$

Attempted inductive definition of the natural numbers (the finite cardinals): the number 0 belongs to the concept F, etc. (§55)

(I) $N^0_x F(x) := \forall x\neg F(x)$;

(II) $N^1_x F(x) := \exists x\forall y(F(x) \wedge F(y) \rightarrow x = y)$;

(III) $N^{n+1}_x F(x) := \exists x(F(x) \wedge N^n_y(F(y) \wedge y \neq x))$.

The tentative contextual definition of the cardinality operator "the number which belongs to the concept Φ" in terms of Hume's Principle (cf. §65)

$N_xF(x) = N_xG(x) := Eq_x(F(x),G(x))$.

The final explicit definition of the cardinality operator (§68):

$N_xF(x) := \#\xi(Eq_x(\xi(x),F(x)))$.

The number which belongs to the concept *F* is the extension of the concept *equinumerous with the concept F*.

Cardinal arithmetic in The Foundations, some further definitions:

n is a cardinal number ($N(n)$) := $\exists F(N_x F(x) = n)$

$0 := N_x(x \neq x)$;

$1 := N_x(x = 0)$;

$\aleph_0 := N_x FN(x)$

n is a finite cardinal number ($FN(n)$) := *n* belongs to the natural series of numbers beginning with 0.

Hume's Principle, taken jointly with the explicit definition of the cardinality operator and the definitions of "0", "1", etc. (note that, according to Frege, each finite cardinal number except 0 can be defined as the number belonging to the concept under which just its predecessors fall) and "$\aleph_1$" provides a means of determining the truth-value of any equation of the following six types (exceptions aside): (1) $N_x F(x) = N_x G(x)$; (2) $N_x F(x) = \#\varepsilon(Eq_x(\varepsilon(x), G(x)))$; (3) $\#\varepsilon(Eq_x(\varepsilon(x), F(x))) = \#\varepsilon(Eq_x(\varepsilon(x), G(x)))$; (4) $N_x F(x) = n$; (5) $n = \#\varepsilon(Eq_x(\varepsilon(x), F(x)))$; (6) $n = m$. An equation of type (4), (5), or (6) can be reduced to one of type (1), (2), or (3), and the truth-conditions of the latter three are determined via the right-hand side of Hume's Principle, or, spelled out more fully, via the one-to-one correlation of the objects falling under $F(x)$ with those falling under $G(x)$.

"Higher" numbers by abstraction in The Foundations

In *Frege 1884*, §104, Frege deals briefly with fractions, irrational and complex numbers. Just as in the case of cardinal numbers, here, too, he write

> everything will in the end depend on the search for a judgeable content which can be transformed into an equation, whose sides are just the new numbers. In other words, we must fix the sense of a recognition-judgement for such numbers. In doing so, we must bear in mind the doubts that we discussed (in §§63-68) with respect to such a transformation. If we follow the same procedure as we did there, then the new numbers will be given to us as extensions of concepts.

In a first step, Frege has to contrive a suitable equivalence relation R_{eq} for the case of, let us say, the real numbers, which can be defined in purely logical vocabulary. In a second step, the real numbers are tentatively introduced by transforming $R_{eq}(\varepsilon, \varepsilon)$ into an identity of real

numbers $\varepsilon(\alpha) = \varepsilon(\beta)$ and by presenting this transformation as a contextual definition of the $\varepsilon\varepsilon$ operator. For the sake of convenience, I refer to the hypothetical abstraction principle for the reals "$\varepsilon(\alpha) = \varepsilon(\beta) \leftrightarrow R_{eq}(\alpha, \beta)$" as "AR".

The domain of the first-order variables in Basic Laws and the referential indeterminacy of course-of-values terms

Assumption "(A)" (to be discussed at length): ***A number of remarks Frege makes in The Basic Laws suggest that he takes the first-order domain to be unrestricted or all-inclusive.***

The permutation argument in *Basic Laws*, §10 (however, in the talk I shall only mention it):

Suppose h is a one-to-one function of all objects of the first-order domain of Frege's logical system. Then the criterion of identity that applies to the references of course-of-values terms, namely the coextensiveness of two functions $\Phi(\xi)$ and $\Psi(\xi)$, also applies to the references of function-value names of the form "h($\dot{\Phi}\Phi(\xi)$)". On this assumption, the following equation is true:

(1) $(\dot{\Phi}\Phi(\xi) = \dot{\Psi}\Psi(\xi)) \leftrightarrow (h(\dot{\Phi}\Phi(\xi)) = h(\dot{\Psi}\Psi(\xi)))$.

From (1) and Axiom V follows

(2) $(h(\dot{\Phi}\Phi(\xi)) = h(\dot{\Psi}\Psi(\xi))) \leftrightarrow \forall x(\Phi(x) \leftrightarrow \Psi(x))$.

Consequently, so Fege argues, Axiom V fails to fix completely the reference of a course-of-values term "$\dot{\Phi}\Phi(\xi)$", at least if there is a bijection h such that for some course-of-values, say $\dot{\Phi}\Phi(\xi)$, $\dot{\Phi}\Phi(\xi) \neq h(\dot{\Phi}\Phi(\xi))$. In more modern terms, Frege's argument can be presented as follows:

> (P1) Suppose $\mathfrak{A}$ is the intended assignment of objects to course-of-values terms satisfying Axiom V. Let h be a non-trivial permutation (of all objects), and consider the assignment $\mathfrak{A}'$ of objects to course-of-values terms which is related to $\mathfrak{A}$ as follows: If Δ is assigned by $\mathfrak{A}$ to a given course-of-values term and $\Gamma = h(\Delta)$, then Γ is assigned by $\mathfrak{A}'$ to that course-of-values term. It follows that $\mathfrak{A}'$ is an assignment of objects to course-of-values terms distinct from $\mathfrak{A}$, but such that it satisfies Axiom V if $\mathfrak{A}$ does.

A special variant of (P1):

> (P2) As in (P1), let $\mathfrak{A}$ be an assignment of objects to course-of-values terms satisfying Axiom V. Let f(ξ) and g(ξ) be two particular, extensionally non-equivalent functions. Let $\mathfrak{A}$ assign *a* to "$\dot{\varepsilon}$f(ε)" and *b* to "$\dot{\varepsilon}$ g(ε)". Let h be a function such that

(i) h(*a*) is the True,

(ii) h(the True) is *a*,

(iii) h(*b*) is the False,

(iv) h(the False) is *b*, and,

(v) for every argument x distinct from these, h(x) = x.

Finally, let □' be an assignment of objects to course-of-values terms related to □ as in (P1), with respect to the particular permutation h just specified. Then, as in (P1), □' will satisfy Axiom V.

In §10, the True and the False are identified with their unit classes by invoking the permutation argument. The True is identified with ⌜(—□) — the horizontal function —□ is elucidated as a concept under which only the True falls; it is obviously coextensive with □= (□= □) — and the False with ⌜(□ = □□*x*(*x* = *x*)) (I use modern notation for Frege's symbols for negation and the universal quantifier).

In one place of my talk, I consider a hypothetical (third primitive) monadic second-level function in Frege's logical system: □ □(□). I refer to it as 'Russell function'.

One example in support of my claim that not only in The Foundations but also in Basic Laws Frege takes the first-order domain to be all-inclusive, at least when he comes to elucidate and define first-level functions. (For reasons of time, I shall only mention the example in my talk).

The instructive example I have in mind is Frege's definition of the "membership-function" (the relation of an object falling within the extension of a concept) □□ □in §34.

a □ *u* := \#*y*[□□*f*(*u* = #*xf*(*x*) □ □ *f*(*a*) = *y*)].

(According to Frege's stipulations concerning the (definite) desciptive function \□ in §11, it holds: If the function □(□) is a concept under which exactly one object falls, then "\□□(□)" refers to that very object. In the remaining three cases — if more than one object falls under □(□); if no object falls under □(□); if □(□) is not a concept — "\□□(□)" has the same reference as "□□(□)".)

Before setting up the definition of □□ □, which involves second-order quantification, Fregestresses that □□ □ must be explained (defined) for all possible objects as arguments. I take the phrase "for all possible objects as arguments" to mean: for all objects whatever as arguments. It cannot plausibly mean: for all courses-of-values as arguments. Here are my

reasons. After having defined $\dot{\varepsilon}\dot{\varepsilon}$, Frege gives a number of explications and closes §34 by summarizing:

> two cases must be distinguished if the value of the function $\dot{\varepsilon}\dot{\varepsilon}$ is to be determined. If the $\dot{\varepsilon}$-argument is a course-of-values, then the value of the function $\dot{\varepsilon}\dot{\varepsilon}$ is the value of that function whose course-of-values is the $\dot{\varepsilon}$-argument for the $\dot{\varepsilon}$-argument as argument. If, on the other hand, the $\dot{\varepsilon}$-argument is not a course-of-values, then the value of the function $\dot{\varepsilon}\dot{\varepsilon}$ is $\Gamma(\dot{\varepsilon} = \dot{\varepsilon})$ for every $\dot{\varepsilon}$-argument [that is, the extension of an empty concept].

Now if Frege had intended to define $\dot{\varepsilon}\dot{\varepsilon}$ only for courses-of-values and the two truth-values qua special courses-of-values as arguments, he could and should have confined himself to stating what the value of $\dot{\varepsilon}\dot{\varepsilon}$ is, if the $\dot{\varepsilon}$-argument is a course-of-values. Yet he does not proceed in this way. "If, on the other hand, the $\dot{\varepsilon}$-argument is not a course-of-values" cannot mean: "if the $\dot{\varepsilon}$-argument is a truth-value". In the light of the identification of the True and the False with their unit classes, this would not make sense, because the value of $\dot{\varepsilon}\dot{\varepsilon}$ for the two truth-values (their unit classes) as $\dot{\varepsilon}$-argument is already determined via the stipulation of the first case: it is the value of that function whose course-of-values is the $\dot{\varepsilon}$-argument for the $\dot{\varepsilon}$-argument as argument. Consequently, nor can it mean "if the $\dot{\varepsilon}$-argument is a truth-value or any other object distinct from both the True and the False and any course-of-values". It can only mean: if the $\dot{\varepsilon}$-argument is any object distinct from any course-of-values.

If for Frege a sound elucidation of "$\Gamma\dot{\varepsilon}(\dot{\varepsilon})$" had been feasible, that is, one which did not rest on a presupposed acquaintance with courses-of-values, then he could have defined straight away the predicate "*a* is a course-of-values" ("*CV*(*a*)"), modelled on his definition of "*n* is a cardinal number" in *Frege 1884*, §72:

$$CV(a) := \Box\dot{\varepsilon}(\dot{\varepsilon}\dot{\varepsilon}(\dot{\varepsilon}) = a).$$

Frege's proof of referentiality in Basic Laws, §31 — the case of the course-of-values operator "$\Gamma\dot{\varepsilon}(\dot{\varepsilon})$"

In *Basic Laws*, §29, Frege states five criteria of referentiality for concept-script names of five different cateories. With respect to "$\Gamma\dot{\varepsilon}(\dot{\varepsilon})$", it suffices to mention the first criterion (A) for monadic first-level function-names, the second (B) for proper names and the fourth (D) for monadic second-level function-names. For the sake of perspicuity, I present these criteria in a formalized version and apply them directly to "$\Gamma\dot{\varepsilon}(\dot{\varepsilon})$". I use "$\Box$"as a semantic predicate

expressing the property of having a reference or of being referential or of referring to something. According to Frege, the proof that "⌜□□(□)⌝" has a reference requires, in addition to the application of criterion (D), an application of criteria (B) and (A). According to (D), we have

□⌜□□(□)⌝ □ □□(□□□(□) □ □⌜□□(□)⌝).

As to the *consequens* — I call names of the form "⌜□□ (□)⌝" *semantic target terms* for "⌜□□(□)⌝" with respect to (D) — Frege must apply criterion (B):

□⌜□□(□)⌝ □ □*X*(□*X*(□□) □ □*X*(⌜□□(□)⌝)) □□□□(□□□(□□, □□) □ □□□(⌜□□(□)⌝, □□) □□□□(□□, ⌜□□(□)⌝)).

□nd to "□□(⌜□□(□)⌝, □□)" and "□□(□□,⌜□□(□)⌝)" — both one-place function-names are semantic target terms for "⌜□□(□)⌝" — criterion (A) applies:

□□□(⌜□□(□)⌝, □□) □ □□□(□□□ □ □□□(⌜□□(□)⌝, □□)).

□□□(□□, ⌜□□(□)⌝) □ □□□(□□□ □ □□□(□□, ⌜□□(□)⌝)).