HOW ABSTRACTION WORKS

LEON HORSTEN and HANNES LEITGEB
University of Bristol

Abstract
In this paper we describe and interpret the formal machinery of abstraction processes in which the domain of abstracta is a subset of the domain of objects from which is abstracted.

1 Abstraction in the History of Philosophy

Plato distinguished the world of ideas or forms from the world of experience. His distinction has been with us ever since. Plato's theory of the orderly world of forms was a first description of what we would now call the abstract realm.

The question about the ontological relation between the abstract realm and the realm of the given proved to be very difficult. One famous bone of contention was the issue of ontological dependence. Plato held that the existence of abstract entities is in some sense independent of the sensory realm. Indeed, he thought that the existence of the world of experience is dependent on the abstract realm: the sensory world consists of "shadows" of ideas. Aristotle riposted that it is rather the other way round. The existence of the forms depends on the sensory objects in which they are realized; the forms only exist in the sensory objects.

Aristotle added that on the cognitive side there is a constructive mental process associated with the relation between forms and objects. The human mind actively abstracts or extracts forms from the objects in which they are realized. This is done by "forgetting" some aspects of objects and focussing on others. Note that Aristotle does not say much about how relations (rather than properties) are abstracted: his was a monadic instead of a polyadic point of view.

A standpoint intermediate between Plato and Aristotle can be adopted. One can hold with Aristotle and against Plato that the existence of an abstraction somehow depends on the given, and at the same time hold with Plato and against Aristotle that an abstraction does not exist *in* the objects from which it is abstracted. In other words, abstract objects are not multiply instantiated—they do not exist in space and time. But their existence depends on the objects from which they are abstracted nonetheless. This is sometimes called a “light” conception of abstract objects.[1] This is the viewpoint that we shall adopt.

Note that we do not wish to imply that all abstract objects can be viewed in this way. Modern set theory, for instance, seems at least on the face of it to postulate many more abstract entities than can stand in some sort of abstraction relation with the given. But, again, the thought that there may be too many abstract objects to leave a direct signature or mark in the given is a modern thought. We confine our attention in this article to the abstracta which are somehow reflected in a fairly direct way in the given.

Another modern observation is that this dependence of abstract objects on the objects from which they are abstracted can be relativized. The given on which an abstractum depends does not have to consist of sensory objects. One can start from concepts, for instance, and abstract from them. Or one can start with mathematical entities, and abstract from them.

2 To Mathematics and Back Again

The idea of abstraction has at some time migrated from philosophy to mathematics. A method of abstraction has been and continues to be fruitfully applied in all areas of mathematics. We are of course talking about the method of introducing new objects by *taking equivalence classes*. You are given a class of mathematical entities G. (These are already abstract entities.) And you are given an equivalence relation R on G. R will partition G into equivalence classes. Each equivalence class can be regarded as an element of a new class of mathematical entities A. The elements of A are regarded as abstracted from G through R. The class of new entities A is disjoint from the old class G and is totally and immediately determined by it through the equivalence relation R.

Examples abound. Here are a few:

Example 1

1. *G is a collection of straight lines; R is the relation of parallelism; A is a collection of directions.*
2. *G is the collection of pairs of integers (with 0 excluded from the second coordinate); R is the relation of being an integral multiple of*

[1] We owe this term to Øystein Linnebo.

(for both nominator and denominator); A is is the set of rational numbers.

Often the new entities in *A* are regarded as constituting a mathematical domain in their own right on which more structure can be defined. Some of this structure is typically "lifted" from the underlying domain *G* to the new domain *A*.

Note that in these cases mathematicians typically do not say that the objects of *A* exist only "in" objects of *G*. So this aspect of the Aristotelian viewpoint is not adopted. But the elements of *A* are in some sense dependent on those of *G*. And, also in accordance with Aristotle, there is a constructive flavour to this. *A* is in a sense *generated* from *G*. Hence the anthropomorphic phraseology.

Frege was one of the first philosophers to realize the importance of the method of taking equivalence classes for philosophy. He, and Carnap after him, sought to apply it in philosophy. Especially Carnap also sought to bring the method of equivalence relations to bear on the empirical realm (in his *Aufbau*). Thus abstraction returned from its mathematical journey to where it was born in the days of Aristotle.

Here are some examples of attempts to apply the method of equivalence classes to philosophy:

Example 2

1. *G is a collection of letter tokens; R is the relation of having the same shape; A is a collection of letters (types).*
2. *G is a collection of sentences; R is the relation of synonymy; A is a collection of meanings (fine-grained propositions).*
3. *G consists of monochromatic colour experiences; R is the relation of perceptual indiscriminability; A is a collection of colour shades.*

In many cases (such as in the last of these examples) it is not clear that the relation *R* that is involved is an equivalence relation: often it is a similarity relation that is not transitive. Carnap was clearly aware of this: he explicitly sought to apply the method of taking equivalence relations to the situation where there is no suitable equivalence relation at hand. We shall leave this development aside here. Instead, we return to Frege.

3 Frege

Frege did not respect the modern requirement that the domains G and A must be disjoint. He explored abstraction mechanisms that require that A Í G. Thus he short-circuited the method of taking equivalence classes. He takes adequacy conditions for such abstraction processes to be given by so-called *level two* abstraction principles. Frege’s two most famous examples are [Frege 1884]:

Example 3

1. *(Law V)* $\{x|Fx\}=\{x|Gx\}\leftrightarrow\forall x{:}Fx\leftrightarrow Gx$
2. *(HP)* $n(F)=n(G)\leftrightarrow F\equiv_{card}G$

Russell showed that Basic Law V is inconsistent. It has been shown that Hume’s Principle (*HP*) is consistent. So the abstraction principles that Frege has in mind are inherently risky.

Abstraction principles regulate identities and differences between presented abstracta. In the method of taking equivalence classes, the abstracta never play a role in evaluating an instantiation of the right-hand-side of the relevant abstraction principle. Thus all identities and differences between presented abstractions are settled in one go. Let us say that an abstractum has been generated when all identities and differences involving presentations of it have been settled. Then the method of taking equivalence classes generates abstracta in one swift movement. Of course we know from the Julius Caesar problem that abstraction principles by themselves do not settle the question what the objects of a given kind are. But given an abstraction principle which involves an equivalence relation on an underlying domain, we have a uniform way of generating abstract entities satisfying that principle: the method of taking equivalence classes.

In Frege’s case of numbers, the identity conditions of presented abstracta involve other identities and differences between presented abstracta. So our question becomes: *can we come up with a general method for generating abstracta when the equivalence relation itself involves the abstract entities already?* There is a fear, because of the circularity, that the process of generating abstracta never gets off the ground. But if an Archimedean starting point can be found, then in some cases all the identities and differences can be settled over the course of a series of stages.

Boolos informally describes Frege's numerical abstraction process in [Boolos 1990, pp. 248f]:

> For how does Frege show that the number 0 is not identical with the number 1? Frege defines the number 0 as the number belonging to the concept *not identical with itself.* He then defines 1 as the number belonging to the concept *identical with 0*. Since no object falls under the former concept, and the number 0 falls under the latter, the two concepts are, by logic, not equinumerous, and hence their numbers are, by Hume's Principle, not identical [...] 2 arises in like manner: Now that 0 and 1 have been defined and shown different, form the concept *identical with 0 or 1*, take its number, call it 2, and observe that the new concept is coextensive with neither of these concepts *because the distinct objects 0 and 1 fall under it*. Conclude by Hume's Principle that 2 is distinct from both 0 and 1.

So there is an ontological dependence of identities and differences between concepted abstracta on identities and differences between other concepted abstracta. Especially the differences are important.

4 Abstraction Models

We will now sketch a general framework for describing the dependency relation that is involved in Fregean self-reflexive abstraction processes, as we may call them. The framework is implicit in [Leitgeb 2005]. We are going to give a somewhat simplified description of the framework. Then we shall illustrate it on the basis of three examples.

Roughly, the idea is this. The identities and differences between some presented abstracta do not depend on identities and differences between presented abstracta at all. These will be our Archimedean starting points. But many identities and differences will only be determined once certain other identities and differences are settled. Thus the identity and difference conditions of presented abstracta can depend on other identities and differences. Identities and differences are no longer settled in one go, but are determined in stages. At some point, this "settling process" gives out.

The objecthood of an abstractum presupposes that the abstractum has been given determinate identity conditions. In Quinean terms: *no entity without identity!* Thus the objecthood of abstracta can depend on the objecthood of other abstracta. (This is of course a thoroughly un-Aristotelian idea.)

If at the end of the process all identities and differences have been settled, then all presentations present abstract objects with an associated determinate identity relation. If not, then matters are less clear.

Let $L_{\equiv}$ be a formal language that contains the relation symbol ≡. We will consider models M^i for the language $L_{\equiv}$ that differ in what they assign to ≡. We shall assume that all nonlogical symbols except ≡ receive the same interpretation in all models M^i. Every model M^i interprets ≡ as an equivalence relation, and for every equivalence relation R there is an M^i that interprets ≡ as R. The domain of the models consists of presentations of abstracta. The aim is to arrive at reasonable answers to questions of the form: *does presentation p present the same abstractum as presentation q?* This will be done by systematically and successively revising the interpretation of ≡. If two presentations stand in the relation ≡ according to a model M^i then the model "judges" these presentations to present the same abstractum; if not, then it judges them to present different abstracta.

Let $\Phi(x,y)$ be a formula that is interpreted as an equivalence relation by all models M^i. The formula $\Phi(x,y)$ which figures on the right-hand of abstraction principles will be the engine for revising the interpretation of °. The ground model M_0 is the model where ≡ is interpreted as the identity relation. This is done in order not to prejudge identity and difference questions.

5 Three Applications

Rather than describing and investigating self-reflexive abstraction processes in general terms, we shall now present three concrete examples of self-reflective abstraction.[2] These examples will convey how such abstraction processes unfold.

5.1 Truth

We let $L_{\equiv}$ be the language of first-order arithmetic plus the equivalence symbol ≡. The arithmetical vocabulary is interpreted in the standard way. Note that this means that this language contains both a symbol standing for *real* identity (=) and a symbol standing for an equivalence relation which does not in all models stand for real identity (≡).

[2] The framework is studied in more generality and detail in [Horsten, Leitgeb, Linnebo in prep.].

The domain consists of codes of sentences, which are regarded as presentations of truth values; we denote codes by using quotation marks.

We consider the following abstraction principle:

$$``\phi"\equiv``\psi"\leftrightarrow(\varphi\leftrightarrow\psi)$$

Call this the *Tarski Abstraction Principle*. Then

$$x\equiv``0=0"$$

can be interpreted as a self-reflexive truth predicate.

5.2 Numbers

We let $L_{\equiv}$ be the language of second-order logic (without the identity symbol) plus the equivalence symbol ≡. We consider only full second-order models.

The domain is taken to consist of codes of open formulas with one free variable, which are regarded as “concepts”. We regard these concepts as presentations of numbers.

We consider the following abstraction principle:

$$``\phi(x)"\equiv``\psi(x)"\leftrightarrow(\varphi(x)\equiv_{card}\psi(x))$$

This is *Hume’s Abstraction Principle*. The right-hand-side expresses the second-order notion of standing in one-to-one onto correspondence. Since we do not have equality in the language, the role of identity in the right-hand-side of Hume’s Principle must be played by the equivalence relation expressed by ≡. In other words, we are *counting* objects using an equivalence relation on the underlying domain that may not coincide with the real notion of identity.

5.3 Events

We let $L_{\equiv}$ be the language of first-order logic (without the identity symbol) plus a causality relation symbol C plus the equivalence symbol ≡.

The domain consists of presentations of events, and C is some sort of similarity relation indicative of causality.

This is our abstraction principle:

$$e\equiv f\leftrightarrow\forall x[(\exists y\colon x\equiv y\wedge C(y,e)\rightarrow\exists y\colon x\equiv y\wedge C(y,f))\wedge$$
$$(\exists y\colon x\equiv y\wedge C(e,y)\rightarrow\exists y\colon x\equiv y\wedge C(f,y))]$$

Let us call this *Davidson's Abstraction Principle*. It says, roughly, that a presentation *e* presents the same event as a presentation *f* if and only if the event(s) presented by *e* and *f* have the same causes and effects.

6 Abstraction Unfolded

The intended interpretation of the equivalence relation symbol $\equiv$ is: *presents the same abstractum*. At the beginning of the process, in order not to prejudge any identities and differences, this interpretation is taken to be the identity relation on the underlying domain of presentations. But this choice will typically judge some presentations to present diffferent abstracta even though in reality they present the same abstractum. So the aim is to improve on this initial choice in stages. This is done in the following way.

An abstraction process is a sequence $\langle E_\alpha\rangle_{\alpha\in On}$ with for each *a*, $E_\alpha = \langle E_\alpha^+, E_\alpha^-\rangle$. E_α^+ contains "settled" identity facts; E_α^- contains "settled" difference facts. The sequence $\langle E_\alpha\rangle_{\alpha\in On}$ is defined by a recursion over the ordinals that we will now describe.

Definition 1 *An equivalence relation E is E_α-respecting if and only if:*

- $E_\alpha^+ \subseteq E$,
- $E \cap E_\alpha^- = \varnothing$.

The idea is that putative identity relations always have to respect the identities and differences that have already been settled in the process. Since the models in an abstraction process only differ by their interpretation of the equivalence symbol *E*, such models can be denoted as M^E. Let *P* be the set of presentations (pairs $\langle d_1,d_2\rangle$ below will be members of $P\times P$).

Definition 2

- $E_\alpha^+ = \varnothing$.
- $E_\alpha^- = \varnothing$.
- $E_{\alpha+1}^+ = \{\langle d_1,d_2\rangle | M^E, \langle d_1,d_2\rangle$ *satisfy* $F(x,y)$ *for all* E_α*-respecting* $E\}$.
- $E_{\alpha+1}^- = \{\langle d_1,d_2\rangle | M^E, \langle d_1,d_2\rangle$ *satisfy* $\neg F(x,y)$ *for all* E_α*-respecting* $E\}$.
- $E_\lambda^+ = \bigcup_{\alpha<\lambda} E_\alpha^+$ *for l a limit ordinal.*
- $E_\lambda^- = \bigcup_{\alpha<\lambda} E_\alpha^-$ *for l a limit ordinal.*

Clearly the successor step in such a process corresponds to van Fraassen’s notion of supervaluation. All abstraction processes are constant from some $l \in On$ onwards. For the first such ordinal l, $E_\lambda = \langle E_\lambda^+, E_\lambda^- \rangle$ is the fixed point of the abstraction process. E_λ consists of the *grounded identities and differences*.

In the case of truth, this process generates (roughly) the least supervaluation fixed point of Kripke’s theory of truth [Leitgeb 2005]. In the case of the numbers, at the first stage the number 0 is differentiated from all other abstracta, at the next stage the number 1 is differentiated from all other abstracta, and so on. In other words, the abstraction process unfolds in exactly the way in which it was described in the passage from [Boolos 1990] that was quoted earlier. In the case of events, it all depends what ground model (collection of event presentations with a causality relation defined on it) we start from. In some cases, all identities and differences will be grounded, in other cases this will not be so [Horsten 2009].

7 Abstract Objects

There are cases in which in the least fixed point all identities and differences are settled. Indeed, this is precisely what happens in the case of Hume’s Principle. In such cases, all identities and differences are grounded. Then E_λ will be an equivalence relation on the domain of presentations of abstracta. And we can simply apply the method of taking equivalence classes, whereby a collection of abstract objects is generated.

But it can also happen that at the least fixed point not all identities and differences are settled. Indeed, this is what happens in the case of truth. In the least fixed point it is determined that the abstractum presented by “*0=0*” is different from the abstractum presented by “*it is true that 0=1*”. But it is not determined whether the abstractum presented by “*0=0*” is numerically identical with the abstractum presented by the liar sentence that says of itself that it is not true.

It is not immediately clear what to say in such cases. For in the case of truth, for instance, this would mean that identity conditions of the truth value presented by “*0=0*” (i.e., the truth value *True*) have not been fully determined. If we interpret the Quinean dictum *no entity without identity* in the strictest terms, then we should probably say that in such cases no realm of abstracta is generated.

Nevertheless, in a weaker sense the Quinean dictum might still be satisfied at least for many cells in E_λ^+. We would like to consider as generated abstracta *some* maximal subsets A of E_λ^+, such that for each a,b of the underlying domain which occur in some identity facts in A, the identity fact $\langle a,b\rangle$ itself belongs to A. It might be necessary to apply the supervaluation idea once more in order to determine such maximal sets.

References

[Boolos 1990] Boolos, G. *The standard of equality of natural numbers.* Reprinted in: William Demopoulos (ed.), *Frege's Philosophy of Mathematics.* Harvard University Press, 1995, pp. 234–254.

[Frege 1884] Frege, G. *Grundlagen der Arithmetik. Eine logisch mathematische Untersuchung über den Begriff der Zahl.* W. Koebner, 1884.

[Horsten 2009] Horsten, L. *Impredicative identity criteria.* Forthcoming in Philosophy and Phenomenological Research.

[Horsten, Leitgeb, Linnebo in prep.] Horsten, L., Leitgeb, H., Linnebo, Ø., *Abstraction, dependence, and groundedness.* In preparation.

[Leitgeb 2005] Leitgeb, H. *What truth depends on.* Journal of Philosophical Logic 34 (2005), pp. 155–192.