On Ordered Pairs

M. Randall Holmes, Boise State

March 29, 2009

The subject of this talk is the ordered pair. We will talk a little about the prehistory of the concept. We will present some ordered pair definitions: Norbert Wiener gave the first one in 1914. We usually use the "simpler" definition due to Kuratowski. Quine gave yet another definition which has advantages in the context of New Foundations but which also has interesting properties in the context of ZFC.

With all these definitions of the ordered pair, what *is* an ordered pair anyway? We discuss ordered pairs as an abstract data type. We give an argument, due to Adrian Mathias, which we think incorrect but interesting, to the effect that abstractness of the ordered pair data type implies the axiom of replacement. We think it is incorrct because Mathias oversimplifies the abstract data type interface for the pair, in view of its actual application in set theory (to the definition of relations and functions).

Can we do without the ordered pair? Zermelo proved his well-ordering theorem in Zermelo set theory without having an ordered pair definition at his disposal; we might find it hard even to state the theorem without pairs! The form of a famous theorem of Sierpinski is conditioned by his use of a pair-free way of representing orders.

In unpublished work, the Belgian mathematician Henrard showed how to define the theory of cardinalities of sets using just three types, without using a notion of pair. This means that NF_3, the fragment of New Foundations whose comprehension scheme consists of all the instances of comprehension which can be typed using three types, has a theory of cardinal number. We have extended his general approach to get a theory of (most) functions in TT_3 (and so in NF_3); we can also show that we cannot get a complete theory of functions in TT_3 by a permutation argument. If the universe can be linearly ordered, it is easy to define a pair in TT_3.

Prehistory of the Ordered Pair I – Russell

Bertrand Russell knew perfectly well that relations and functions should be thought of as sets of ordered pairs ("couples"). However, in *Principia Mathematica* he and Whitehead are hampered by the lack of a way to define ordered pairs in terms of sets. As a result, the type system of *PM* contains a type of n-ary relations for each sequence of n types (actually, these types are further subdivided into "orders", a complication which we will not visit here).

Ordered pairs (couples) *are* defined in *PM* – as a species of relation! The pair (a, b) is the relation which holds between x and y iff $x = a \wedge y = b$.

It is reported that Russell responded with no particular interest when Norbert Wiener informed him of his pair definition which made it possible to collapse the type theory of *PM* to a simple linear hierarchy (as long as one also followed Ramsey in abandoning the orders).

Prehistory of the Ordered Pair II – Zermelo

Zermelo's axioms of set theory of 1908 provide a theory which can found all of classical mathematics. But Zermelo, like Russell, did not know how to represent pairs as sets. Nonetheless, he gives definitions and proves theorems in his 1908 paper which we would have difficulty stating, much less proving, without the use of ordered pairs.

For example, he defines $A \sim B$ (for A and B disjoint) as holding iff there is a set F such that for each $a \in A$ there is exactly one b such that $\{a, b\} \in F$ and for each $b \in B$ there is exactly one $a \in A$ such that $\{a, b\} \in F$.

In his proof of the Well-Ordering Theorem he represents partial well-orderings of a set using the collection of their initial segments (this representation of well-orderings will reappear in this talk).

The First Definition – Wiener

Norbert Wiener defined the pair (x, y) as $\{\{\{x\}, \emptyset\}, \{\{y$ in 1914. This pair is adequate for all mathematical purposes in either type theory or Zermelo set theory. In *PM* as we have noted it makes it possible to work with a simple linear hierarchy of pairs, since relation types are no longer needed. If a and b are of type i, (a, b) will be of type $i + 3$, and for any sets A and B of type $i + 1$, $A \times B$ will exist in type $i + 4$, and any relation R with domain A and range B will be coded in type $i + 4$ as a subset of $A \times B$.

It is remarkable that Russell was apparently entirely unimpressed by this simplification of his theory.

In Zermelo set theory, the axiom of elementary sets (which provides $\emptyset$, singletons, and unordered pairs) ensures that Wiener pairs exist.

For any sets A and B, the set $A \times B$ is a subset of $\mathcal{P}^3(A \cup B)$, so it exists by application of the Zermelo axioms of separation, power set, union, and elementary sets (pairing). Any relation R with domain A and range B will be definable as a subset of $A \times B$ by separation.

The Usual Definition – Kuratowski

Later, Kuratowski introduced the definition which is now usual, $(x, y) = \{\{x\}, \{x, y\}\}$.

The advantage of this definition over Wiener's is that (in the context of type theory) the pair (x, y) lives two types above x and y rather than three. The cartesian product of $A \times B$ is obtained from the second iterated power set of $A \cup B$ rather than the third (basically the same phenomenon).

The *disadvantage* is that the proof of the basic property of the ordered pair (we highlight its statement below because it is important) is significantly harder.

Baic Property of the Ordered Pair: $(a, b) = (c, d) \leftrightarrow a = c \wedge b = d$

For the Wiener pair this is very easy. Suppose $\{\{\{x\},\emptyset\},\{\{y\}\}\} = \{\{\{z\},\emptyset\},\{\{w\}\}\}$. There is exactly one element of (x, y) which has two elements ($\{\{x\},\emptyset\}$), exactly one of which is a singleton, whose sole element is x. The same argument shows that this uniquely determined object is also z, so $x = z$. There is exactly one element of (x, y) which is a double singleton, and the sole element of its sole element is y, and of course also w, so $y = w$.

We have all seen the argument for the Kuratowski pair, which is complicated by the fact that the two formal elements of the Kuratowski pair (x, y) are in fact the same object when $x = y$, which adds nasty case analysis to the proof.

Why doesn't this matter?

But we don't care. We use the Kuratowski pair.

The reason for this is that we don't care what definition of the ordered pair we use, as long as it satisfies the Basic Property [we will see below that this is something of an oversimplification]. Once we have proved that the pair has the basic property, we have no occasion [little occasion?] to look at the internal details of the definition.

Ordered pairs make up what computer scientists call an *abstract data type*: the pair construction, whatever it is, needs to satisfy the Basic Property, and as long as it does, it will serve our mathematical purposes [we will see below that the abstract data type interface actually requires *a little* more information about the properties of the pair].

Another Ordered Pair – the Quine Pair

The one clear advantage that the Kuratowski pair has over the Wiener pair is most briefly expressed in terms of type theory: the Kuratowski pair is two types higher than its projections, whereas the Weiner pair is three types higher.

Willard Quine pointed out that in the presence of Infinity there is a pair which is the same type as its projections. This is most useful in the context of type theory or New Foundations, in which it makes it much easier to talk naturally about cartesian products and multiplication (for example), but we will see that the Quine pair has at least one appealing property in the context of the usual set theory ZFC.

The definition of the Quine pair is quite baroque; we give it in full excruciating detail, just because we think it is fun.

We suppose the natural numbers implemented. We define $\sigma(x)$ as $x+1$ if x is a natural number and x otherwise. We define $\sigma_1(x)$ as $\sigma``x$, the elementwise image of $x$ under $\sigma$. We define $\sigma_2(x)$ as $\sigma``x \cup \{0\}$. It is useful to note that for any x and y, $x = y \leftrightarrow \sigma_i(x) = \sigma_i(y)$ for $i = 1, 2$, and $\sigma_1(x) \neq \sigma_2(y)$ for any x and y.

Now we define (a, b) as $\sigma_1``a \cup \sigma_2``b$. This works as a pair because it is the union of disjoint sets $\sigma_1``a$ and $\sigma_2``b$ from which a and b can be recovered. It is further the case that *every* set is a Quine ordered pair.

In type theory with infinity, as long as a and b are of type at least 4, (a, b) is defined and of the same type. In Zermelo set theory or New Foundations (a, b) exists for each a, b, and every set is an ordered pair.

The interesting property which this pair has in the context of the usual set theory is that for any infinite ordinal α, $V_\alpha \times V_\alpha = V_\alpha$. The ordinal rank of the Quine pair (a, b) is the maximum of the ranks of a and b, while the rank of the Kuratowski pair is two higher.

This feature might have some applications in presentations of the structure of L, for example.

A one-type differential – Holmes

I exhibit a pair of my own creation which is one type higher than its projections and which does not require existence of an infinite set.

Let 0,1,2,3,4,5,6,7 be eight distinct objects (whose precise nature is unimportant).

$(x, y) = \{(x', 0, 1), (x', 2, 3), (y', 4, 5), (y', 6, 7) \mid x' \in x \wedge y' \in y\}$

An argument of Mathias

Adrian Mathias has argued that abstractness of the ordered pair datatype implies the Axiom of Replacement. His argument is interesting but I think it actually demonstrates that the abstract data type interface of the ordered pair contains a little more information than just the Basic Property.

Suppose that $(\forall x \in A.(\exists! y, \phi(x, y)))$.

Define an ordered pair $\langle a, b\rangle$ as $(0, (a, b))$ if $a \notin A$ and as $(c, (a, b))$ where c is the unique object such that $\phi(a, c)$ holds if $a \in A$. This pair clearly satisfies the Basic Property [it may not be exactly Mathias's pseudo-pair (I'm working from memory), but it captures his idea]. So by abstractness we can rely on cartesian products to exist: we write $A \otimes B$ for $\{\langle a, b\rangle \mid a \in A \wedge b \in B\}$. The set

$$\{x \in \bigcup{}^2(A \otimes A) \mid (\exists y \in A \times A.(x, y) \in A \otimes A)\}$$

(note that the pair here is the usual one and that both kinds of cartesian product appear) can also be written

$$\{y \mid (\exists x \in A.\phi(x, y))\},$$

establishing that Replacement holds!

Our critique of this argument has to do with the *purpose* of the pair. We introduce the pair not for its own sake but to support the theory of relations and functions. For the pair to successfully support the theory of relations and functions (with domain A and range B, for example), we need the cartesian product $A \times B$ to exist. This is of course the property of the usual pair that Mathias illicitly exports from his weird pair to the usual pair in the above argument to establish Replacement. We note that another property is also required in order to be able to define domains and ranges of sets: any set R of ordered pairs needs to be a subset of some cartesian product $A \times B$ (whence we can define $\texttt{dom}(R)$, for example as $\{x \in A \mid (\exists y.(x, y) \in R)\}$. Notice that the fact that an appropriate $A \times B$ in the case of the Kuratowski pair is $\bigcup^2 R \times \bigcup^2 R$ is used in the argument above. These assertions about sets of pairs are relevant to the purpose of the pair and show that the ordered pair data type, properly understood, is implemented in Zermelo set theory without Replacement.

Can we get along without the ordered pair?

We noted above that Zermelo managed to do quite a lot in his 1908 paper without using pairs at all. Do we actually need to implement pairs to get a theory of functions and relations?

Interesting work along these lines, which has never been published, was done by the Belgian mathematician Henrard, and I have recently extended this to some extent. Henrard showed that it is possible to define a complete theory of cardinality in the theory of types with just three types (and so in the fragment NF_3 of New Foundations defined and shown to be consistent by Grishin in 1969, an untyped set theory in which only the instances of comprehension which are typable using three types are used). This definition is perfectly usable in ordinary set theory, and one can prove such results as the Schröder-Bernstein theorem in the three type context. Note that in type theory with three types the Kuratowski pair is definable but useless, as one can define the Kuratowski pair only of type 0 objects, which will be type 2 and not capable of being a member of a set.

We follow the lines of our extension of Henrard's concept (our definition of cardinality is quite different from his, though it is in the same spirit). The starting point is the observation (also exploited by Zermelo in his 1908 paper and by Sierpinski to reduce by one the height of the tower of exponentials in his theorem $\aleph(\kappa) < exp^3(\kappa)$ of choice-free mathematics) that a partial order (on type 0) can be represented by the (type 2) collection of the (type 1) segments of the partial order.

Coding Relations

Suppose that $x\,R\,y$ is notation for a transitive, reflexive relation (a quasi-order). Instead of capturing R using a collection of ordered pairs, capture it as a collection of "segments". Define x_R as $\{y \mid y\,R\,x\}$ and define R as the set $\{x_R \mid x \in \mathtt{dom}(R)\}$.

If R is a set which codes a relation in this way, then x_R is the intersection of all elements of R which contain x.

For R any set of sets at all, we define $x\,R\,y$ as $(\forall A \in R.x \in A \rightarrow y \in A)$. The relation R defined in this way is reflexive and transitive. Of course the same quasi-order can now be coded by more than one set. A condition which picks out the originally intended relation codes is $R = \{x_R \mid x \in \bigcup R\}$. Each quasi-order has a unique code satisfying this condition.

Transitive reflexive relations (quasi-orders) include two important categories of relation as special cases: equivalence relations, which are represented by the associated partition, and partial orders, which are represented by the sets of their closed segments (and also by the set of all their segments).

It is amusing to observe that the Kuratowski ordered pair (x, y) is the code for the minimal quasi-order $\leq$ such that $x \leq y$. This has the flavor of the *PM* definition of the ordered pair.

Functions as Quasi-Orders

Any function f determines a quasi-order: we define $x \geq_f y$ as holding iff $(\exists n.y = f^n(x))$. Alternatively, we can define the segment in $\geq_f$ determined by x as the intersection of all sets which contain x and are closed under f. This is also called the "forward orbit" of f.

For any $x \in \mathtt{dom}(f)$, we define x_f as $(\exists n.y = f^n(x))$, and we code f by the set $\{x_f \mid x \in \mathtt{dom}(f)\}$. It is worth noting that x_f and the set we have just constructed as code for f are definable in type theory with three types (or in ordinary set theory) if the function f is defined by a formula ($y = f(x)$ equivalent to a formula $\phi(x, y)$). In type theory x and y would be type 0, x_f would be type 1 and f would be type 2.

In most but not all cases we can define $f(x)$ using the element x and the code f for the function.

If $\{x\} \in f$, then $f(x) = x$.

The ideal condition is that $f(x)$ is the immediate successor of x in the quasi-order: if there is a unique y such that $y \neq x$, $x \notin y_f$, and $x_f = y_f \cup \{x\}$, then this y is $f(x)$.

If the forward orbit of x has two elements, then for some $y \neq x$, $x_f = \{x, y\} = y_f$, and in this case $y = f(x)$. Notice that this same thing occurs if $f(x) \notin \texttt{dom}(x)$: there is no way to distinguish this case from the case of a 2-cycle, so we extend all functions f to map elements of $\texttt{rng}(f) - \texttt{dom}(f)$ back to their preimages, and make sure to specify intended domains in applications.

The bad case in which we cannot for the moment recover $f(x)$ is that in which the forward orbit x_f is finite and $f(x)$ (and possibly x itself) belongs to a finite cycle with an element other than x or $f(x)$.

We now give a partial formal definition of function.

For any set f and element x of $\bigcup f$, we define x_f as the intersection of all elements of f which contain x.

A *nice function* is a set of sets f with the following properties:

1. For all $x \in \bigcup f$, $x_f \in f$.

2. Every element of f is of the form x_f for some element of f.

3. Every element x_f is either of the form $\{x\} \cup y_f$ for a uniquely determined $y \neq x$, or has one element, or has two elements.

We define $f(x)$, where f is a nice function and $x \in \bigcup f$, as x in case $x_f = \{x\}$, as the other element of x_f if $|x_f| = 2$, and as the unique $y \neq x$ such that $x_f = \{x\} \cup y_f$ otherwise.

Cardinality

Suppose $A \sim B$. This means there is a one-to-one function from A onto B. Let f be such a function. Define the set coding f as above. A problem will exist if there are finite cycles in the function f, which will be finite sets x_f with more than two elements (finite cycles of order 2 with additional preimages also cause problems but cannot occur in this context). Replace each such element of the set coding f with the set of singletons of its elements to obtain a set f^* which codes a nice function which differs from f in fixing each element of each finite cycle in f. Since each element of a finite cycle belongs to $A \cap B$, it is easy to see that f^* still codes a bijection from A to B.

So we can define $A \sim B$ as holding iff there is a nice function f with the property that for each $x, y \in A$, $f(x) = f(y) \rightarrow x = y$. Notice that the fact that elements of $B - A$ are mapped back into A in a way which might violate injectivity is ignored.

We can prove basic properties of cardinality using this definition, such as reflexivity, symmetry and transitivity of equinumerousness, and the Schröder-Bernstein theorem. An odd thing which happens is that if f witnesses $A \sim B$ and g witnesses $B \sim C$, then $A \sim C$ is not necessarily witnessed by $g \circ f$, because $g \circ f$ is not necesssarily a nice function: however, we can define $(g \circ f)^*$ as above and show that it works.

More Functions

We describe a device for constructing a code for a general function which might have finite cycles. Let f be a function (perhaps defined by a formula). Define the code for f as above. If there are no large cycles (finite orbits with more than one member), this will be our final code for f. Otherwise (and on an additional assumption) we add more elements to f to obtain a code from which we can define $f(x)$ in all cases (with the usual extension to $\mathtt{rng}(f) - \mathtt{dom}(f)$). The additional assumption is that for each large cycle $c \in f$ we can select an element x_c. This would be true if $\mathtt{dom}(f)$ were linearly orderable, for example.

The additional elements we add are all subsets of large cycles. For each x such that $c \subseteq x_f$, where c is a large cycle, we add $x_{f'}$ as a new element of the code for f, where f' is the function which differs from f only in fixing each x_c and each $(f\lceil c)^{-1}(x_c)$. x'_f is an orbit modified by being truncated at any x_c or at the unique preimage of an x_c in c.

In the extended code of f, the cycles are identifiable as the only elements of f which are disjoint unions of two elements of f (c is the union of $(x_c)_{f'} = \{x_c\}$ and $(f(x_c))_{f'} = c - \{x_c\}$), and the special elements x_c are identifiable as the sole element of the singleton element of the 2-partition of a cycle. Removing the cycles from the extended code for f gives the code for f': we can then define $f'(x)$ as above, and define $f(x)$ as mapping each x_c to the only element of c which is not an image under f', and as mapping each $(f\lceil c)^{-1}(x)$ (identifiable as the only element of c other than x_c which is fixed by f') to x_c. There is a special case where c is the union of two singletons, so we cannot specify x_c; in this case f maps each of the two elements of c to the other.

Definition of General Function

A general function is a set of sets f with the following properties.

1. For each $x \in \bigcup f$, $x_f \in f$.

2. Any element of f which is not of the form x_f is the union of two disjoint sets belonging to f, one set a singleton and the other coding a finite linear order. Any two distinct elements of f not of the form x_f are disjoint. The set obtained by removing all the sets not of the form x_f from f is a nice function f'.

We define $f(x)$ as follows. For each x not belonging to an element of $f - f'$, we define $f(x)$ as $f'(x)$. In the case where both elements of the partition of $c \in f - f'$ are singletons, if the two elements of c are x and y, we define $f(x) = y$ and $f(y) = x$. If $c \in f - f'$ has more than two elements, we define x_c as the element of the singleton element of the unique partition of c into two elements of f. We define $f(x_c)$ as the unique element of c which is not the image under f' of an element of c (the minimum element under $\geq_{f'}$ of the linear order element of the partition). Define y_c as the other fixed point under f' in c (the maximum element under $\geq_{f'}$ of the linear order element of the partition); we define $f(y_c)$ as x_c. For each other element z of c, we define $f(z)$ as $f'(z)$.

This only works to define functions in all cases if we have choice from disjoint collections of finite sets. Further, it is possible to show by a Frankel-Mostowski permutation argument that there are functions with 3-cycles which cannot be represented by sets at all in the theory of types with three types, in a suitable precise sense.

If we have a uniform method of choosing one element from any unordered pair (as we would have if there were a linear order of the universe, for example) then there is a sensible ordered pair which can be used to define relations and functions in the usual way in type theory with three types. Choose two disjoint sets A and B each of size 5. Define (a, b) as $A\Delta\{a, b\}$ if a is the selected member of $\{a, b\}$ and as $B\Delta\{a, b\}$ otherwise. This limits the sphere in which the pair-free theory is likely to be considered. But it does have applications:

for example, the theory of order type (which requires isomorphisms between well-orderings, which can be represented in our scheme in a choice-free context since they are partial orders) is handled by this representation of functions because the domain of an isomorphism between well-orderings is well-ordered.

There can be no reliable notion of function in TT_3

Consider a model of TT_4 (just add one more type) in which there is a function f (represented by a set of Kuratowski pairs as usual) whose domain is an infinite union of cycles of length 3.

We use a permutation method to modify this model. Consider all elements A of the model (at any positive type) which have the property that for some finite subset B of the domain of f, A is invariant under all permutations of type 0 which are obtained by iterating f 0-2 times independently on each cycle in f which is a subset of $V - B$. It is straightforward to show that these sets (the sets of finite support with respect to a certain group of permutations of type 0) make up a model of TT_4. This is essentially the Frankel-Mostowski method for showing that Choice is independent of *ZF* with atoms.

Note that f has finite support (in fact, empty support) with respect to the group, so it still exists in the model. But the new model cannot contain any choice set for the finite orbits under f, since all but finitely many of the orbits must contain 0 or 3 of the elements of any given set (because all but finitely many of the orbits are outside the support of the set).

Thus we cannot expect to code f as above. But we can say more: we cannot code f at all in the first three types. For any set X of type 1, we have already seen that the elements of all but finitely many orbits have exactly the same relations to X; similarly, for any type 2 set X and all but finitely many orbits $\{a, b, c\}$, and for any Y of type 1, either all of $\{a, b\} \cup Y$, $\{a, c\} \cup Y \{b, c\} \cup Y$ are in X or none of them are. In fact, the sets with finite support in our group in types 1-2 are the same as the sets with finite support in the larger group containing all permutations of each of the orbits independently. This means that there can be no 3-typed formula $H(x, y, f^*)$ equivalent to $y = f(x)$, since the pair $\langle x, y \rangle$ will have exactly the same relations to the parameter f^* that the pair $\langle y, x \rangle$ does in all but finitely many cases where $y = f(x)$ is true. But the formula $y = f(x)$ (though it involves 4 types) can be used as a "black box" in otherwise 3-typed formulas to define sets (since f does have finite support and we still have a model of TT_4. If we had a function definition which worked for all 3-typed formulas, it would be expected to work to define f in this case as well.