Similarity Inheritance: A New Model of Inheritance for Spreadsheet VPLs*

Rebecca Walpole Djang and Margaret M. Burnett
Department of Computer Science
Oregon State University, Corvallis, Oregon 97331
{walpolr, burnett}@cs.orst.edu

Abstract

Although spreadsheets can be argued to be the most widely-used visual programming languages (VPLs) today, most are very limited compared to other VPLs, supporting only a few built-in types and offering only primitive support for code reuse. The inheritance mechanisms of object-oriented programming might seem to offer help for the latter problem, but incorporating these mechanisms in a traditional way would introduce concepts foreign to spreadsheets, such as message passing. In this paper, we present similarity inheritance, a new approach to inheritance that is suitable for seamless integration into the spreadsheet paradigm. We first explain the model independently of any implementation, and then present a prototype implementation in the research spreadsheet VPL Forms/3. We show that bringing inheritance functionality to the spreadsheet paradigm can be done using the widely-understood idea of copy/paste. Further, we show why the approach requires the presence of a live, visual environment.

1. Introduction

Spreadsheets have proven to be a popular programming paradigm, accessible even to non-programmers. Current spreadsheets, however, suffer from some of the problems that have been solved in other programming languages. For example, in other programming languages, object-oriented inheritance mechanisms have improved upon ad-hoc (cut-and-paste) reuse of code, but spreadsheets still support only ad-hoc reuse through copy/paste and formula replication. Thus spreadsheet users must remember the reuse relationships themselves and maintain them manually whenever a reused formula changes. Some commercial spreadsheets such as Excel® have a few additional conveniences, such as automated formula adjustment when a new copy of a linked spreadsheet is made. However, these features are simply editing conveniences, and the user is still left to manually maintain the reuse relationships.

It occurred to us that incorporating inheritance into spreadsheets could result in stronger support for formula reuse than is found in current spreadsheets. However, existing models of inheritance do not seem suitable for the spreadsheet paradigm because they introduce concepts foreign to spreadsheets, such as message passing. Thus, we set out to find an approach to inheritance suitable for spreadsheet VPLs.

We use the term spreadsheet VPLs to refer to a variety of systems that follow the spreadsheet paradigm, from commercial spreadsheets to more sophisticated research VPLs that follow the declarative, one-way constraint evaluation model. The essence of the paradigm is summarized by Alan Kay's value rule for spreadsheets [Kay 1984], which states that a cell's value is defined solely by the declarative formula explicitly given it by the user.

In this paper we present a new approach to inheritance suitable for spreadsheet VPLs, and an instantiation of the approach in the research spreadsheet VPL Forms/3 [Burnett and Gottfried 1998; Atwood et al. 1996]. The approach, called similarity inheritance, provides a concrete way of sharing behavior among objects in a spreadsheet VPL. The unique attributes of similarity inheritance are that:

- it includes an explicit visual representation of all the object's unique and shared behaviors, rather than leaving some behaviors implied through parenthood;
- it is flexible enough to allow sharing at multiple granularities and even allows mutual inheritance;
- it brings object-oriented concepts to spreadsheet VPLs without using external languages or macros;
- it subsumes the current spreadsheet edit-based mechanisms for formula propagation, unifying formula reuse with inheritance.

*This work was supported in part by Hewlett-Packard, by Harlequin, by the National Science Foundation under grant CCR-9308649 and an NSF Young Investigator Award, and by a NASA Graduate Student Researcher Award.
2. Related work

2.1 Alternatives to traditional inheritance

The most prevalent alternative to the traditional class/subclass model of inheritance has been prototype-based inheritance. The prototype model is more concrete than the class/subclass model because inheritance in prototype-based languages is based on concrete parent objects rather than abstract classes. If an object cannot handle a message directly, it delegates it to its parent object, which in turn handles it or delegates it to its parent, and so on. Prototypes remove the need for the concepts of class, subclass and instance since any object can be used as the basis for defining a new object. Self [Ungar et al. 1991] is perhaps the best-known prototype-based language. ObjectWorld [Penz 1991] is a prototype-based language that, like ours, uses visual mechanisms to emphasize concreteness. However, unlike our approach, ObjectWorld does not use inheritance, instead achieving code reuse through object composition.

An important difference between our similarity inheritance approach and most prototype-based languages is that our model does not use message passing. Kevo [Taivalsaari 1993] is one of the few prototype-based languages that does not use message passing. Kevo is a textual language that emphasizes the concreteness and self-sufficiency of objects. Operations can be marked as applying to individual objects or to clone families which are groups of similar objects automatically inferred by the system. Thus Kevo does not require a designated parent prototype for a collection of objects, but consequently there are no change propagation mechanisms for objects outside the clone family.

Most approaches to inheritance operate at the granularity of entire classes or objects. Mixins [Bracha and Cook 1990] are a technique for providing inheritance of partial classes. Also known as abstract subclasses, mixins exist only to be inherited by other, complete classes. They usually define just a small piece of functionality and, combined with multiple inheritance, can cut down on the code duplication that arises when the language allows inheritance only at the class level. Another approach to fine-grained inheritance is found in the language 1+ [Ng and Luk 1995]. 1+ inheritance is not determined by subclassing, but by explicitly listing the methods to inherit. Neither of these approaches supports a granularity finer than whole methods.

2.2 Combining spreadsheets with object-oriented programming

Many VPLs have incorporated object-oriented concepts [Cox et al. 1989, Burnett et al. 1995, Braine and Clack 1997]. However, spreadsheets and more advanced spreadsheet VPLs have had little work to date on approaches to inheritance, perhaps because there has been only a little work that incorporates support for objects into spreadsheets.

Commercial spreadsheets provide support only for a few built-in types—numbers, Booleans, and strings—as first-class values, and do not provide a formula-based mechanism allowing users to add new types of objects. Although some spreadsheets gain partial support for additional objects through the use of macro languages and “trapdoors” to other programming languages (such as Visual Basic), these approaches do not maintain a seamless integration with the spreadsheet paradigm, because they use notions such as global variables, state modification, and imperative commands in a language different from the formula language of the spreadsheet.

A few research spreadsheet VPLs have also incorporated external languages to support object-oriented features. ASP (Analytic Spreadsheet Package) is a spreadsheet VPL in which every cell can be any object, and every formula is written in Smalltalk [Piersol 1986]. Smedley, Cox, and Byrne have incorporated the VPL Prograph with user interface objects into a conventional spreadsheet for GUI programming [Smedley et al. 1996]. C32 [Myers 1991] is a spreadsheet VPL that is part of the Garnet and Amulet user interface development environments [Myers et al. 1990; Myers et al. 1996]. C32 does not itself feature the graphical creation and manipulation of objects; these must be created by other tools or written in Lisp or C++. These approaches add some of the power of object-oriented programming, but do not enforce consistency with the value rule, since global variables and state-modifying mechanisms circumvent it.

Some research spreadsheet VPLs have moved toward expanding the types of objects supported without the use of external programming languages. A pioneering system in this direction was NoPumpG [Lewis 1990] and its successor NoPumpII [Wilde and Lewis 1990], two spreadsheet VPLs designed to support interactive graphics. These languages include some built-in graphical types that may be instantiated using cells and formulas, and support limited (built-in) manipulations for these objects, but do not support user-defined objects.

Penguins [Hudson 1994] is a spreadsheet VPL for specifying user interfaces. Penguins supports composition of objects by collecting cells together, and formula inheritance at the object level. Unlike our work, it employs several techniques that do not conform to the spreadsheet value rule, such as interactors objects that can modify the formulas of other cells, and imperative code similar to macros.
3. Background: Introduction to Forms/3

We have created a prototype implementation of our approach to inheritance in the spreadsheet VPL Forms/3, and the examples in this paper are presented in that language. This section provides the background in Forms/3 necessary to understand the examples.

A Forms/3 programmer creates a program by using direct manipulation to place cells on forms (spreadsheets) and to define a formula for each cell using a flexible combination of pointing, typing, and gesturing. A program's calculations are entirely determined by these formulas.

Forms/3 has long supported an extensible collection of types. Attributes of a type are defined by formulas in cells, and an instance of a type is the value of a cell, which can be referenced just like any cell. For example, the built-in circle object shown in Figure 1 is defined by cells defining its radius, line thickness, color, and other attributes. One way to instantiate a circle is to copy the circle form, changing any formulas necessary to achieve the desired attributes (as in the figure); another way is to graphically define its attributes [Burnett and Gottfried 1998], such as by sketching a new circle or by stretching an existing circle by direct manipulation. The graphical method is a shortcut for the first method, and we will use only the first method in this paper.

To implement a new user-defined type of object, the Forms/3 programmer provides cells and formulas to construct a prototypical object. During the construction the system responds, as one would expect on a spreadsheet, with immediate visual feedback as each new formula is entered. The formulas specify the internal composition of the object, how it should appear visually on the screen, and any operations that it provides.

The internal composition is defined by cells and matrices that can be placed inside abstraction boxes, which provide information hiding. Abstraction boxes are cells whose formulas default to being the composition of their components. For example, Figure 2 shows a stack implemented by a one-dimensional matrix (inside abstraction box Stack), in which the programmer has placed the sample value “hi”. Because cell Stack has a sample value, as soon as the formula for cell top is

![Figure 1: The white primitiveCircle form is a built-in form that defines a prototypical instance of type primitiveCircle. The gray 392-primitiveCircle form is a copy that has been modified to describe a different instance. The circle in cell newCircle is defined by the other cells, which specify its attributes. To refer to the circle elsewhere in the program, a formula can reference 392-primitiveCircle:newCircle. The programmer cannot view the formula (primitive implementation) of newCircle, but can view and specify the other cells' formulas by clicking on their formula tabs. Radio buttons and popup menus (e.g., lineForeColor) provide a way to reliably enter constant formulas when only a limited set of constants are valid.](image1)

![Figure 2: (Top): The user's view of Stack hides the internal implementation and displays stacks using the formula the programmer has provided for a distinguished Image cell. (Bottom): The stack implementor's view of object Stack with most of the cell formulas visible. (Matrices in Forms/3 are not required to be homogeneous.)](image2)
entered, top displays Stack's top element ("hi"). The other formulas are also programmed in this concrete way; they reference Stack and immediately display their own results based on the sample. The sample values (in cells with visible formula tabs) on copies of form Stack can be replaced by references to other cells in the program, which provides the functionality of incoming parameters in traditional languages.

4. Similarity inheritance model

In this section we show how the preceding approach to objects in spreadsheets can be extended to support inheritance. We begin by describing our new model of inheritance, independently of any language implementation. In the model description, we will use object-oriented terminology to facilitate comparison with other models of inheritance, although it will later be demonstrated (Section 5.5) that the approach is not restricted to relationships among objects, and can be used for relationships among Excel-like spreadsheets as well.

We define the similarity inheritance model to be comprised of a model of interaction (between the programmer and the computer) and a semantic model. The interaction model is defined by the tuple:

\[(\chi, \delta, \lambda, \rho)\]

where \(\chi\) is the copy operation that creates a shared definition, \(\delta\) is the formula definition operation, \(\lambda\) is a liveness level 3 or higher from Tanimoto's liveness scale [Tanimoto 1990] indicating that immediate semantic feedback is automatically provided\(^1\), and \(\rho\) is a representation mechanism that explicitly includes all shared formulas and relationships in each object's representation.

Three important points about the interaction model are: (1) it separates the syntax with which the human communicates to the computer about program semantics from that used by the computer to communicate to the human about program semantics (for example allowing animations or graphical views), (2) it does not necessarily map to a static textual syntax (for example, it allows dynamic syntaxes) and (3) it depends on environmental characteristics that are not usually guaranteed by textual programming languages but are common in visual ones. Note that the elements of the interaction model are not mere editing details of an environment, but rather define the general characteristics upon which our semantics rest.

The semantic model can now be defined as follows. Each object \(O\) in a program is a set of formula definitions \(\{O_d1,O_d2,...,O_dn\}\). For example, in Forms/3, an object is a form. Each \(O_dj\) is a formula residing in a cell. The symbol \(\rightarrow\) (pronounced "shares with") indicates a shared definition; the arrow points from the original version to the copied one. The semantics of \(\rightarrow\) are

\[A_{di}\rightarrow B_{dj} \Rightarrow A_{di}=B_{dj}\]

The operations \(\chi\) and \(\delta\) determine when \(\rightarrow\) holds, as summarized in the following table (\(i \in \{1...n\}\)):

<table>
<thead>
<tr>
<th>Operation</th>
<th>Precondition</th>
<th>Postcondition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\chi) applied to (A) (B) does not exist &amp; (B) is a new object &amp; (\forall i, A_{di}\rightarrow B_{dj})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\chi) applied to (A) and (B) are existing objects, (A\neq B) &amp; (A_{dk}\rightarrow B_{dk})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\delta) applied to (A) (B) is an existing object &amp; (\forall A, A_{dk}\rightarrow B_{dk})</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first row defines large-grained similarity, and means that if \(\chi\) is applied to an object \(A\), a similarity relationship will be created between \(A\) and a new object \(B\) such that all of \(A\)'s definitions share with \(B\) (this can be abbreviated \(A\rightarrow B\)). The second row defines fine-grained similarity, which allows a single definition \(A_{di}\) to be copied to object \(B\) to create a shared relationship between \(A_{di}\) and \(B_{dj}\). The third row implies that overriding removes any "upstream" sharing relationships, but not "downstream" relationships.

Due to element \(\rho\) of the interaction model, objects in the similarity inheritance model have the property of self sufficiency from the programmer's perspective, meaning that every supported operation for an object and every piece of data it contains can be determined by examining the object itself rather than also requiring the inspection of parent objects or descriptive classes. The implication of the \(\lambda\) element of the model is that the programmer creates and manipulates live objects while constructing the program, rather than abstract descriptions of objects.

From this model, differences between similarity inheritance and other approaches become clear. For example, the class-based model has a \(\rightarrow\) relationship defined between classes (not objects). The prototype-based model supports \(\rightarrow\) between objects, and both the prototype-based model and the class-based model support \(\rightarrow\) at the granularity of methods (overriding), but neither supports \(\rightarrow\) at the granularity of methods. Similarity inheritance is also different from both models in that it allows not only multiple inheritance but also mutual inheritance. Multiple inheritance occurs in cases such as \(A\rightarrow B, C_{d1}\rightarrow B_{d1}, C_{d2}\rightarrow B_{d2}\) and \(C_{d3}\rightarrow B_{d3}\). Mutual inheritance occurs in cases such as \(B_{d2}\rightarrow C_{d2}\) and \(C_{d3}\rightarrow B_{d3}\).
5. Similarity inheritance in Forms/3

5.1 Interaction model

The interaction model is instantiated in the research spreadsheet VPL Forms/3 as follows. Operation $\chi$ is supported by a copy form button, which copies the form selected in a scrolling list, and by a paste button on each form, which pastes selected cells onto the form. Recall that $\chi$ is not simply an edit operation, but defines the $\rightarrow$ relationship. Operation $\delta$ is supported by allowing the programmer to edit any formula that is visible. Liveness level $\lambda$ is level 4, so after every formula edit, immediate visual feedback is given about the edit's effect on the program. In Forms/3's representation $\rho$, each definition (cell and formula), whether copied or not, is visible, which allows it to be edited by operation $\delta$. Shading indicates whether a form or cell is copied. Section 6 explains additional features of Forms/3’s representation.

5.2 Large-grained inheritance

The Stack in Section 3 is an example of an object created from scratch. Multiple Stack objects can be created with the copy operation $\chi$, as was illustrated by the circle example. But suppose the programmer wants not another stack but something else that is similar to a stack, for example, a queue. Taking advantage of that similarity, the programmer can start with a copy of Stack, then modify the behavior using operation $\delta$. Inheritance and overriding of cell names as well as of cell formulas are allowed. A change to the push operation and some renaming of cells are all that is required to turn the copy into a Queue (see Figure 3).

Note that using similarity inheritance, the programmer identifies that two objects are simply similar in implementation or purpose. In contrast to this, in a class-based language, the programmer may spend extra time wondering what the “right” relationship is between a stack and a queue. One is not a subtype of the other, yet they are similar. In fact, extra work to reorder the inheritance hierarchy may be needed in some cases just to add one new class. As we saw above, however, similarity inheritance allows the programmer to create a similarity relationship without implying “is-a” subtype or subclass relationships. Instead, it defines a “like-a” relationship. For example, a queue is like a stack except that new items are inserted at the opposite end.

Because of the $\rightarrow$ relationship between Stack and Queue, changes to formula definitions on Stack will propagate to Queue unless they have been overridden. For example, a fix to the push operation on Stack would not have any effect on Queue, but a fix to the pop operation would propagate to the dequeue on the Queue form. Most prototype-based languages lose this ability to propagate changes to groups of objects because of their emphasis on object individuality; instead, shared parts must be abstracted out of the objects, as in Self's traits [Ungar et al. 1991].

5.3 Fine-grained and multiple inheritance

As noted in the discussion of the model, the combination of large-grained and fine-grained similarity allows multiple inheritance. For example, in Forms/3 suppose a new form Deque is created via large-grained similarity from Queue. (A deque is a double-ended queue.) This new object needs to allow items to be added to either end of the queue. The programmer may notice that Stack’s push is exactly the required behavior for Deque and can use fine-grained similarity, copying push, to allow Deque to inherit just that one operation from Stack.

Because of the interaction element $\rho$, the programmer now sees the new cells as part of the definition of Deque also (Figure 4).

Multiple inheritance in other languages can lead to conflicts when more than one method of the same name are unintentionally inherited. By providing inheritance on the level of cells, the similarity model allows the programmer to select only the operations that are actually needed, avoiding unintentional inheritance. (If the programmer does accidentally attempt to introduce a conflict, the system provides options for resolving it at the time of the edit.)
5.4 Mutual inheritance

Suppose, as in Figure 5, someone added the new operations size and empty? to Queue. Another programmer might find those operations useful for Stack as well and copy them to the Stack form. Stack and Queue now both inherit from each other. Like multiple inheritance, mutual inheritance is not a new concept in the language, but rather a feature of the flexibility of similarity inheritance, which makes mutual inheritance straightforward. To the best of our knowledge, similarity inheritance is the first model to support mutual inheritance.

5.5 An end-user example

In the previous sections, we have discussed our approach from the standpoint of how it can be used to share behavior among objects. However, as has been noted earlier, the approach is general enough to allow sharing of other pieces of programs, even when there is no relationship among the types of objects involved. This allows the same approach to be used for simple formula reuse as for object inheritance, instead of prior approaches, which relied on copy/paste and "replicate" options. The advantage of using inheritance for reusing spreadsheet formulas is that the relationships among originals and copies are maintained, supporting automatic propagation of bug fixes and explicit depiction of relationships.

For example, consider Figure 6, which shows a spreadsheet (written in Forms/3) to compute course grades. Suppose the user teaches several sections of the course, and keeps each section in a separate spreadsheet for convenience.

There are two reuse situations in this example: the reuse of the formula for the top row down through the remaining rows of this section, and the reuse of these formulas in other sections. In the first case, traditional spreadsheets use a "replicate" mechanism (copy down the rows). Our system does not apply inheritance to this case; instead, like some other spreadsheets, it has a way to group cells with a common formula. It is the second case in which we apply similarity inheritance. In the second case, traditional spreadsheets use a "copy/paste" mechanism (copying into other sections), and then if the weights need to be changed, the user would have to remember to do all of the copy/pasting again. However, if a system implements copying using similarity inheritance to make the relationships explicit, as does Forms/3, then a change to the weights in the first section can automatically propagate to all the other students.

As this example demonstrates, similarity inheritance can be used not only to maintain relationships among objects, types, and operations, but also among pieces of any sort of calculation. An attractive feature of this generality is that it affords a gradual migration path for users to move from using only simple numbers and strings in their formulas to using more complex objects with inheritance as they gain expertise, since the same mechanism for inheritance is employed for reusing formulas in both situations.
6. Explicit representation

The interaction model requires the existence of an explicit representation, $p$. To design the representation for similarity inheritance in Forms/3, we used a set of design benchmarks [Yang et al. 1997] that are a concrete application of several of the cognitive dimensions for programming systems [Green and Petre 1996]. To focus on the reasons behind our design choices for $p$, we present the representation from the perspective of the benchmarks that had the greatest influence on the design.

The visibility of dependencies benchmark is the ratio of program dependencies that are explicitly visible to the programmer. Green and Petre noted hidden dependencies as a severe source of difficulty in understanding programs. The dependencies of interest in the similarity inheritance model are those created by the $\rightarrow$ relationship. The dependencies can be at either the form or cell level. For each, the programmer may be interested in both "what affects this?" and "what does this affect?" for a total of four kinds of dependencies. Because of the importance of explicit representation to the similarity inheritance model, it was important that all four kinds of dependencies be explicitly represented.

The two "what affects this" questions are handled by legends. A legend under each formula lists the cell that it was directly copied from. If that cell was in turn copied from another, ellipsis follow and the name of the original cell is also given. The same legend mechanism is used for explicit representation at the form level (see Figure 7).

The two "what does this affect" questions are answered by copy dependency arrows (Figure 7) and by a summary view (Figure 8). The summary view represents each form as a node labeled with its name. Arrows between nodes make different kinds of relationships explicit. Our current design has three kinds of arrows indicating form copies, dataflow, and cell copies. The summary view is part of a package of "live" views we previously implemented that address other, non-inheritance oriented, reuse issues [Walpole and Burnett 1997].

Visibility of program logic is another benchmark that influenced our design. The program logic of interest here is the logic of inherited formula. The current design makes an inherited formula visible in the place where it is inherited. This avoids the "yo-yo problem" in class-based languages where to see the program logic for a subclass, the programmer may need to visit several classes up and down the class hierarchy. The yo-yo problem is also exhibited in most prototype-based languages, where instead of class definitions, the programmer must examine multiple levels of parent objects to view inherited code.

Whenever features are added to a VPL's representation, the limited size of the screen must be taken into account. Our representation is designed to fit into a small portion of the programmer's screen. The display of form-level inheritance takes up just one line on the form, no matter how long the list becomes, because intermediate forms are normally elided. Likewise, cell-level inheritance legends take only one line per formula. These legends can also be hidden to conserve screen real estate when needed.

7. Discussion and future work

The similarity inheritance model was devised as a way to bring inheritance particularly to spreadsheet VPLs, but it may be suitable for other kinds of interactive VPLs as well, provided that they support the interaction model. We do not expect the approach to be used in strictly textual languages, which are usually defined apart from any environment, because this characteristic would seem to prevent any guarantee that the required elements of the interaction model would be present.

An important next step will be empirical work to learn whether users make fewer reuse errors or reuse formulas more often under similarity inheritance. Two other questions that we would like to explore are whether users are as comfortable with similarity inheritance as with copy/paste/replicate, and whether the explicit representation succeeds at making the flexibility inherent
in the approach manageable. Finally, we would like to gather empirical data about whether and how people use mutual inheritance.

8. Conclusion

In this paper, we have presented a new model of inheritance for spreadsheet VPLs. The model supports large-grained inheritance, fine-grained inheritance, multiple inheritance, and mutual inheritance. The prototype implementation shows that the model can be incorporated into the spreadsheet paradigm, using only cells and declarative formulas, without violating the value rule or requiring users to learn other programming languages or macro languages.

We have shown also that the approach to inheritance can be used to improve the way reuse relationships among cells are managed even in simple formula reuse, which has been traditionally supported only by copying or replicating a formula to other cells. This flexibility not only increases the support for this kind of operation, it also allows for a gradual path for a user to progress from simple formula copy/paste to more advanced applications of the technique such as inheritance among user-defined types.

An important feature of similarity inheritance is object self-sufficiency. Self-sufficiency enables a concrete style of programming, and avoids the yo-yo problem. These features are a result of the semantic model's reliance on an interaction model. Although two of the four elements in the interaction model—the copy operation and the definition operation—could be accomplished textually, the other two elements—an explicit representation and a liveness level of 3 or higher—require a visual environment. Hence, it is unlikely that similarity inheritance would be possible for a textual language.

References


