The Choose-Your-Own-Adventure Calculus (Pearl/Brave New Idea)

Tomas Petricek
Charles University, Prague, Czechia

Jan Liam Verter
Charles University, Prague, Czechia

Mikoláš Fromm
Charles University, Prague, Czechia

Abstract

TODO: Some of the most remarkable results in mathematics reveal connections between different branches of the discipline. The aim of this paper is to point out a modest, but still remarkable, similarity between a range of different interactive programming systems.

TODO: Say somewhere that this also formalizes things like mixed-initiative interaction and programming by demonstration

TODO: It is mainly about giving us a nice way to talk about lots of things in interactive programming systems and enable a transfer of ideas between them

2012 ACM Subject Classification Software and its engineering → General programming languages

Keywords and phrases Interactive programming systems, Type providers, Proof assistants

1 Introduction

Multiple interactive programming systems, ranging from code editors for object-oriented programming languages to data exploration systems and interactive proof assistants, exhibit a remarkably similar pattern of interaction. They offer the user, who can be a programmer, a data scientist or a proof writer, a range of choices that the user can select from in order to complete their program, script or proof. The user can initiate the interaction iteratively, using it to create and refine a larger part of their program.

There are subtle differences between different implementations of the general pattern. In some systems, the resulting source code will contain a trace of the choices made by the user. For example, when choosing an item from a list of class members, the code will contain the member name. In some systems, the interaction results in a block of code that can be included in the source file, but does not include a trace of the interaction. For example, invoking a proof search or case split in Idris [7] constructs a well-typed program, but leaves no trace of the command used to construct it. The nature of the generated options also varies. The list of choices may include all possible options that are valid at a given location, or it may list only a subset of the valid options. In some cases, it may also include incorrect options as, for example, in auto-completion for dynamic languages [11].

The aim of this paper is paper is to formally capture the recurring interaction pattern:

1. We motivate the formalism by reviewing four different systems that implement a variation on the interaction pattern. These include type providers for data access in F# [45], type providers for data exploration in The Gamma [33, 30], AI assistants for semi-automated data wrangling [36] and tooling for interactive proof assistants [3, 7, 46] (Section 2).
2. We introduce the *choose-your-own-adventure calculus*, which is a small formal structure that models an interactive system where a user constructs a program by repeatedly choosing from a list of options offered by the system (Section 3).
3. The calculus allows us to make the aforementioned subtle differences precise. We define the notions of *correctness* and *completeness* for the choose-your-own-adventure calculus. To distinguish the different ways of embedding the interactions in the edited programs, we also formally define *internal* and *external* mode of system integration.
4. We show that various programmer assistance tools, such as search and AI-based recommendations can be built on top of the primitives offered by the calculus, showing how the choose-your-own-adventure calculus supports of transfer of ideas across different kinds of interactive programming systems.

The main contribution of this paper is conceptual rather than technical. We capture a pattern that is perhaps not surprising in retrospect, but that is easy to overlook until it is given a name. We use formal programming language theory methods to precisely describe interesting aspects of the pattern. Moreover, our work also confirms that programming language theory methods can be extremely effective for studying not just *programming languages*, but also interactive *programming systems* [16].

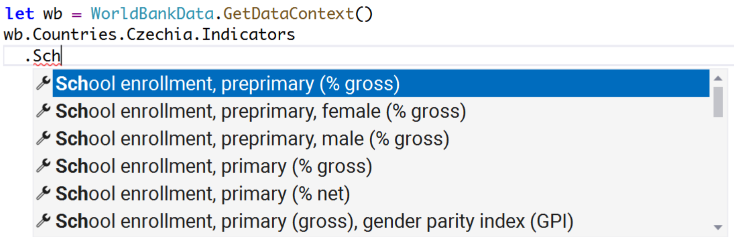

Figure 1 F# code editor showing completions offered by the World Bank type provider.

2 Motivation

Computer scientists studying programming have long focused on programming languages as syntactic entities, sometimes neglecting the interactive environments in which they are inevitably embedded [12]. Notably, in many of the motivating examples that we draw from in this section, the interactive aspect of the system is only described in supplementary materials [7, 45, 3]. Only recently, programming language theory started to be used to study interactive environments [1, 24]. Our work contributes to this research direction.

The following sections review four different instances of the choose-your-own-adventure interaction pattern. In all of those, an interactive editor offers the user some kind of a completion list during working with the system.

Type providers. F# type providers [45] are a mechanism for integrating external data sources into the F# type system. A type provider is a compiler extension, loaded and executed at compile-time and at edit-time. It can run arbitrary code to read the structure of external data and use it to generate a suitable statically-typed representation of the data, typically as objects with members. Type providers can, for example, infer the type from a sample JSON [35] or read a database schema.

The example in Figure 1 shows a simple type provider for accessing information from the World Development Indicators database. The provided wb object allows the programmer to access any indicator of any country in the database by choosing an appropriate [Country] and an [Indicator] in a chain of members wb.Countries.[Country].Indicator.[Indicator]. The result is a time series with values for the given indicator and a country. More generally, the example can be seen as a special case of a type provider for slicing n-dimensional data cube [33] – we choose a fixed value for two of the three dimensions (country, indicator, time).

When using the type provider, the user types the first line of code and triggers auto-completion by typing wb followed by the dot. The rest of the code is constructed by choosing an option from a list and typing another dot.[1]

[1] This interaction pattern has been lightheartedly called *dot-driven development* by Phil Trelford [39].

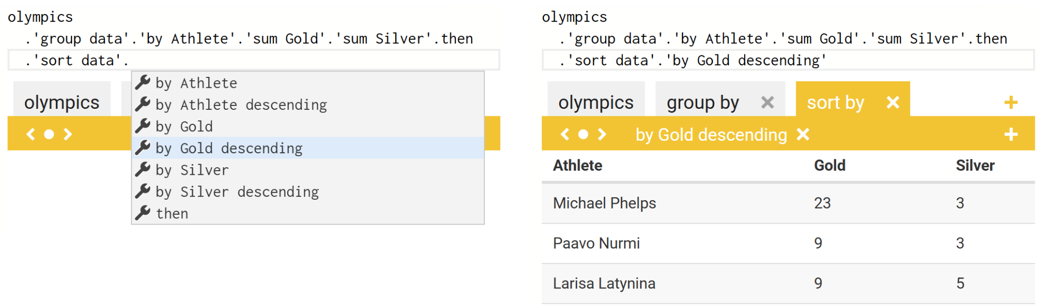

Figure 2 Constructing a query in The Gamma. We count the number of gold and silver medals for each athlete and sort the data by the number of gold medals.

Data exploration. The Gamma [33] is a programmatic data exploration environment for non-programmers. In The Gamma, type providers are the primary programming mechanism. They are used not just for data access, but also for constructing queries.

The type provider shown in Figure 2 lets the user construct an SQL-like query by repeatedly choosing operations and their parameters [30]. It keeps track of the schema and uses it to generate all possible valid parameters. When sortin data, it generates an object with two members for each columns – one for ascending and one for descending sort. Similarly, the grouping operation first offers all columns as possible grouping keys and then lets the user choose from a range of pre-defined aggregations (sum, count, average, concatenate). The system also evaluates the query on the fly, providing a live preview during editing [32].

The interaction pattern is the same as before. After the usertriggers auto-completion, they repeatedly select an operation and its parameters to construct a query. One notable difference is that the structure of the generated types is potentially infinite (the user can keep adding further operations) and so the types are generated lazily.

AI assistants. The third instance of the choose-your-own-adventure interaction pattern comes from the work on semi-automatic data wrangling tools known as AI assistants [36]. An AI assistant guides the analyst through a data wrangling problem such as reconciling mismatched datasets, filling missing values or inferring data format and types. An AI assistant solves the problem automatically and suggests an initial data transformation, but it also generates a number of constraints that the user can choose from to refine the initial solution. If the initial solution is not correct, the user chooses a constraint and the AI assistant runs again, suggesting a new data transformation that respects the constraint.

Figure 3 shows an example. It uses the datadiff [43] AI assistant, running in a Wrattler notebook [34], to merge broadband quality data published by Ofcom for two subsequent years. The format of the CSV files for the two years differs. Columns were added, removed, renamed and their order has changed. In the example, we selected 6 columns from the year 2015 and want to find matching data from 2014.

When the AI assistants runs automatically, it correctly maps the numerical columns, but it incorrectly maps the Urban.rural (2014) column to Nation (2015). This happens because both columns are categorical and have three values with similar distribution. A data analyst can easily spot the mistake. They click the "+" button to add a constraint and choose Don't match 'Urban.rural' and 'Nation' to specify that the two columns should not be matched. Datadiff then runs again and finds the correct matching.

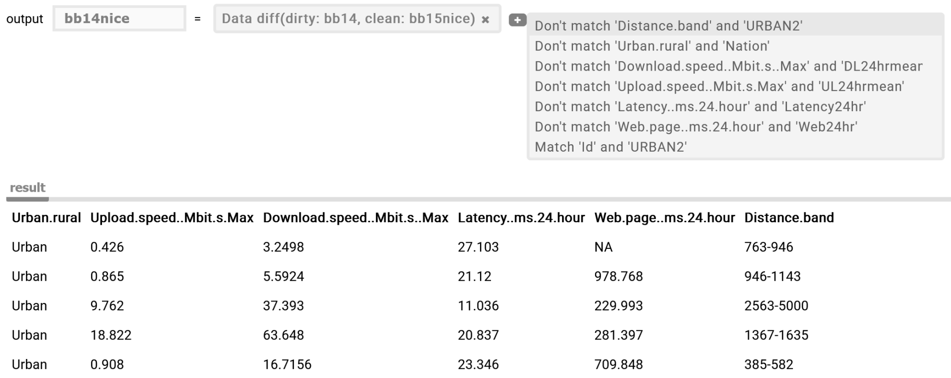

Figure 3 Using the datadiff AI assistant to reconcile the structure of the two datasets. The user is offered a list of constraints to prevent or force matching between specific columns.

The interaction patter is the same as in the previous two cases. The analyst constructs the correct data transformation by repeatedly choosing from a list of options, until they obtain the desired result. However, the way the interaction pattern is implemented differs. First, in the case of type providers, we are gradually constructing a program by adding operations to a method chain. Now, the AI assistant synthesizes a data transformation (program) and we are gradually adding constraints to control the synthesis. Second, in the case of type providers, the completion list offered all possible members of the object. Now, the list offers constraints recommended by the AI assistant which may not be complete.

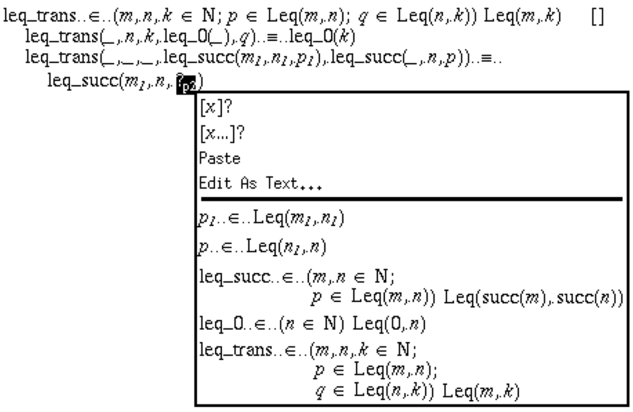

Figure 4 Constructing a proof of the transitivity of the $\leq$ relation in the ALF editor. The user is offered a range of variables and constructors in scope at the current location. [3]

Interactive theorem provers. A fourth example of the choose-your-own-adventure interactive pattern can be found in interactive theorem provers. When writing programs in systems like Idris [8], the user typically works by stating the desired conclusion and filling the implementation with a hole. The system provides a range of interactive editing capabilities to fill the holes [25]. It can, for example, generate a case split or search for a proof [7].

Systems like Idris provide key bindings to invoke the completions, but the functionality could also be offered through a user interface. An example that illustrates this is the interactive editor for the ALF theorem prover [21], which is based on the refinement of an incomplete proof object [3]. This is illustrated in Figure 4. The user is proving the transitivity of the $\leq$ relation for Peano arithmetic natural numbers. They pattern match on the proof argument p and complete the first branch. For the second branch, they need to fill a hole $?_{p2}$ (called a wildcard in ALF). They trigger a completion and a pop-up menu shows the available variables and constructors, including leq_trans that can be used to complete the proof. After choosing leq_trans, two new holes are generated for its arguments. Those can be, again, filled interactively, by choosing p_1 and p from the completion.

The interaction pattern is again the same. The user repeatedly triggers a completion and uses it to refine and complete their proof by filling holes. There are subtle differences too. Unlike with AI assistants, each completion directly refines the proof that the user is editing. Unlike with type providers, a completion may generate multiple new holes, rather than just adding to a chain of operations.

The ALF editor is a historical example, but a similar user interface could be built for systems like Idris or Coq. The two would work differently. As in ALF, Idris source code represents the proof itself and a completion would replace a hole with a suggested term. In Coq, the proof is a series of tactic invocations and so selected completions would be added to this list and would form a trace of the interaction with the user.

3 Formal model

A system that implements the choose-your-own-adventure interaction pattern repeatedly offers the user a range of options to choose from. Each of the options is designated by an identifier. The system also maintains a state during the process which determines subsequent options. The state may not be visible to the user, but the user can always explicitly request the program constructed so far.

We can think of the interaction with the system as navigating through a tree structure, starting from a root and choosing one of the possible branches in each step.[2] In the following definition, the key choices operation can thus be seen as returning branches of a given node.

▶ **Definition 1** (Choose-your-own-adventure system). *Given expressions $e \in \mathbb{E}$ and states $\sigma \in \Sigma$, a choose-your-own-adventure system is a pair of operations* choices, choose *such that:*

- $\mathsf{choices}(\sigma) = \{\iota_1 \mapsto \sigma_1, \ldots, \iota_n \mapsto \sigma_n\}$ *is an operation that takes a state and generates options designated by an identifier ι_i and represented by a state σ_i,*
- $\mathsf{choose}(\sigma) = e$ *is an operation that returns generated program for a given state.*

The definition is not a programming language calculus in the usual sense in that it does not define a concrete syntax with reduction rules. It is an abstract algebraic structure that captures the structure of a system that supports the choose-your-own-adventure interaction pattern. The definition is close to that of an AI assistant [36], which is written using a language specific for the data wrangling domain (such as cleaning scripts or input and output data) but is structurally similar. It is also worth noting that the definition may describe not just trees, but also graphs with cycles – a system can return to an already visited state. This is not practically useful, but it does not pose a theoretical problem.

External mode of embedding. One of the subtle questions about the choose-your-own-adventure pattern raised in the introduction concerns the different ways in which a trace of the interaction is embedded in the interactively constructed program. In the *external mode*, the interaction results in code that becomes a part of the edited program, but it is not possible to reconstruct the steps used to generate the code.

The choose-your-own-adventure interaction pattern is typically used to complete a partial program. To model this, we assume that the host language has a notion of a hole, written as ? and that a user can select a part of program to invoke the completion on. We write $E[e]$ for a completion context, akin to evaluation contexts in operational semantics.

We assume that, for a program containing a hole in a completion context $E[?]$, we can construct an initial choose-your-own-adventure state using an operation $\mathsf{init}(E[?]) = \sigma_0$.

▶ **Definition 2.** *An expression $E[?]$ is completed as $E[e]$ via external embedding of an interaction with a choose-your-own-adventure system consisting of* choices *and* choose *if:*

1. $\mathsf{init}(E[?]) = \sigma_0$ *obtains the initial state of a choose-your-own-interaction system,*
2. σ_n *is a system state such that* $\forall i \in 1 \ldots n \,.\, (\iota_i \mapsto \sigma_i) \in \mathsf{choices}(\sigma_{n-1})$*, i.e., the user makes a series of choices resulting in a final state of the system σ_n,*
3. $E[e]$ *where* $e = \mathsf{choose}(\sigma_n)$*, i.e., the final program is constructed by replacing the hole in the completion context with the expression e generated from σ_n.*

[2] Serving as another evidence for the surprising effectiveness of the concept of a tree [26].

As we will see when we revisit the earlier examples formally, the external mode of embedding is used, for example, in the case of interactive theorem provers like ALF or Idris. In those systems, the user triggers the completion on a proof (program) containing a hole. They then fill the hole and, possibly iteratively, further holes in the generated proof. The final expression is embedded in the source code, but it does not indicate what options, identified by $\iota_1, \ldots, \iota_n$, were selected in the process.

Internal mode of embedding. In the *internal mode*, a trace of the interaction with a choose-your-own-adventure system is embedded directly in the constructed program. This is the case with type providers, where a user chooses a sequence of object members to be accessed. The same would be the case in a completion system for Coq that would offer tactics to apply, becuase the resulting proof would contain a record of the selected tactics.

To talk about the internal mode formally, we again need the init operation, but also an operation decode that extracts identifiers of invoked completions from an expression. An internal embedding is the same as external embedding with an additional constraint:

▶ **Definition 3.** *An expression $E[?]$ is completed as $E[e]$ via internal embedding of an interaction with a choose-your-own-adventure system consisting of* $\mathsf{choices}$ *and* choose *if:*

1. *$E[?]$ is completed as $E[e]$ via external embedding using* init*,* $\mathsf{choices}$ *and* choose *through a series of choices designated by identifiers $\iota_1, \ldots, \iota_n$,*
2. *it also holds that $\mathsf{decode}(e) = (\iota_1, \ldots, \iota_n)$.*

If a choose-your-own-adventure system is integrated in a programming language through internal embedding, we can reconstruct the choices through which the user constructed an expression e in a completion context $E[e]$, assuming they used the interactive system rather than entering the code directly. This also means that we can reconstruct the final state σ_n of the system by starting from $\mathsf{init}(E[?])$ and following the choices specified by $\iota_1, \ldots, \iota_n$.

4 Examples

We now revisit the four examples from Section 2 and show how they fit the above formal model. All four examples rely on some domain-specific logic. We describe what information the logic provides, but do not model it formally. This has been done elsewhere, in works describing the individual systems.

To show how the model lets us distinguish subtle details of interactive programming systems, we start with a model of data exploration system that is inspired by The Gamma, but differs in one notable way. We then discuss type providers more generally and show how to correctly model The Gamma. We then revisit the remaining two examples.

4.1 Data exploration

In The Gamma, the choose-your-own-adventure interaction pattern is used to construct a query that transforms the given input data. The query is a sequence of operations with parameters, $op(p_1, \ldots, p_n)$, loosely modelled after relational algebra [9].

In The Gamma, the query is hidden from the data analyst. Behind the scenes, the system generates objects with members and the identifiers designating individual options are the names of those members. The operation is encapsulated in the code of the accessor of the member. In the simplified model in this section we ignore this fact. The model presented here directly generates code that calls the underlying operations. For example, assume that the user makes the following choices:

«group data» . «by Athlete» . «sum Gold» . «count all» . «then»

In The Gamma, the individual identifiers become object members and they are included as a member chain in the generated code. In the following simplified model, the completion instead fills the hole with an expression representing the operation:

`group("Athlete", sum("Gold"), count())`

The two approaches have different human-computer interaction trade-offs. In terms of cognitive dimensions [13, 6], the latter has a greater closeness of mapping, while the former is less cognitively demanding to read for a non-programmer. As discussed in Section 5, the two implementations of the choose-your-own-adventure interaction pattern also differ in terms of their formal properties.

Formal model. The options generated by The Gamma let the user select both the next operation and the parameters of the previously selected operation. The available operations and parameters are generated based on a schema S that is transformed by the operations. The state of the system σ contains the current schema S and the operations applied so far. In the following, we write $op(\boldsymbol{p})$ for an operation with a vector of parameters:

$$\sigma = S, [op_1(\boldsymbol{p_1}), \ldots, op_n(\boldsymbol{p_n})]$$

The behaviour of the choices operation depends on whether the last operation in the sequence expects further parameters or whether it is fully-specified. In the first case, the recommendation engine generates possible additional parameter values $p', p'', \ldots$ based on the schema S, the operation op_n and the already known parameters $\boldsymbol{p_n}$. The choices operation then generates options that add the additional parameter. We generated the identifiers $\iota', \iota'', \ldots$ based on the state and the parameter value, such as «by Gold descending». Note that adding a parameter may also result in a new schema $S', S'', \ldots$ (which the recommendation engine computes based on the previous schema and the new parameter):

$$\begin{array}{l} \mathsf{choices}(S, [op_1(\boldsymbol{p_1}), \ldots, op_n(\boldsymbol{p_n})]) = \\ \quad \{\ \iota' \mapsto (S', [op_1(\boldsymbol{p_1}), \ldots, op_n(\boldsymbol{p_n}, p')]), \\ \quad\ \ \iota'' \mapsto (S'', [op_1(\boldsymbol{p_1}), \ldots, op_n(\boldsymbol{p_n}, p'')]),\ \ldots\ \} \end{array}$$

If the last operation takes no further parameters, the system produces a choice of possible next operations $op', op'', \ldots$. Again, we are also given new schemas $S', S'', \ldots$ and we generate identifiers $\iota', \iota'', \ldots$ based on the operation name. The choices operation then returns options that add the additional operation:

$$\begin{array}{l} \mathsf{choices}(S, (op_1(\boldsymbol{p_1}), \ldots, op_n(\boldsymbol{p_n}))) = \\ \quad \{\ \iota' \mapsto (S', [op_1(\boldsymbol{p_1}), \ldots, op_n(\boldsymbol{p_n}), op'()]), \\ \quad\ \ \iota'' \mapsto (S'', [op_1(\boldsymbol{p_1}), \ldots, op_n(\boldsymbol{p_n}), op''()]),\ \ldots\ \} \end{array}$$

Finally, the choose operation takes the state σ and generates an expression that represents the data transformation. This is only possible if all parameters are fully-specified. For simplicity, assume that k is the index of the last fully-specified operation (either n or $n-1$). If the host language lets us compose functions using $f \circ g$, we can write:

$$\mathsf{choose}(S, (op_1(\boldsymbol{p_1}), \ldots, op_n(\boldsymbol{p_n}))) = op_1(\boldsymbol{p_1}) \circ \ldots \circ op_k(\boldsymbol{p_k})$$

The recommendation engine behind The Gamma provides a domain-specific logic for generating possible operations and their parameters based on the current schema of the data. As the above definition shows, this underlying engine can be easily exposed through the common choose-your-own-adventure interface.

4.2 Type providers

The type provider mechanism in F# operates at the level of the type system. It is not a merely an editor feature. A type provider for a data source, such as the World Development Indicators database, generates a collection of types that model the external data source. In F#, the types are classes with members that implement the logic to retrieve data at runtime.

The auto-completion mechanism in F# code editors, which implements the choose-your-own-adventure interaction pattern, is not specific to type providers. It offers a list of members of an object based on its type. We model the completion as an iterative process, repeatedly adding further members to a chain. The state σ thus consists of an initial expression on which the completion is invoked, a chain of selected members and the type of the last member.

To model the completion mechanism, we also need to model information about types. We loosely follow the Foo calculus model [35] and write $\mathbb{C}$ for a set of class definitions, each consisting of an implicit class constructor and a collection of members M:

$$\begin{array}{rcl} \sigma & = & e.\iota_1.[\ldots].\iota_n, C \\ \mathbb{C} & = & \{C \mapsto \mathsf{type}\ C(\overline{x:\tau}) = \overline{M},\ \ldots\ \} \\ M & = & \mathsf{member}\ \iota\!:\!C = e \end{array}$$

Each member in the Foo calculus consists of a of a name ι, return type C and implementation e. For our purposes, we only need the type information and so the operations that define the choose-your-own-adventure are parameterized by the set of classes $\mathbb{C}$.

The $\mathsf{choices}_{\mathbb{C}}$ operation finds the class definition corresponding to the type of the last member in the current chain. It offers choices appending each of the available members to the current chain. The $\mathsf{choose}_{\mathbb{C}}$ operation returns the constructed member chain:

$$
\begin{aligned}
&\mathsf{choices}_{\mathbb{C}}(e.\iota_1.[\ldots].\iota_n, C) = \\
&\qquad \{\ \iota' \mapsto (e.\iota_1.[\ldots].\iota_n.\iota', C') \\
&\qquad\ \ \iota'' \mapsto (e.\iota_1.[\ldots].\iota_n.\iota'', C''),\ \ldots\ \} \\
&\quad \text{where } \mathbb{C}(C) \ = \ \mathsf{type}\ C(\overline{x:\tau}) = M', M'', \ldots \\
&\qquad\quad \text{and } M' \ = \ \mathsf{member}\ \iota' : C' = e'
\end{aligned}
$$

$$\mathsf{choose}_{\mathbb{C}}(e.\iota_1.[\ldots].\iota_n, C) = e.\iota_1.[\ldots].\iota_n$$

The model does not directly refer to type providers. Those are responsible solely for generating the type definitions in $\mathbb{C}$ as documented in earlier work [35]. It is worth noting that the type provider for data exploration, implemented by The Gamma, additionally needs to generate classes lazily [30]. To model this aspect, the simple lookup $\mathbb{C}(C)$ needs to be replaced with an operation that returns the type definition, alongside with a new context $\mathbb{C}'$ that contains additional generated type definitions (return types for all the members of the class C).

The model follows the internal mode of embedding the interaction in the program. It is easy to define the decode operation that takes the resulting generated expression and returns the sequence of choices, because the choices are items of the member chain. A slight caveat is that the completion is not invoked on an empty hole, but on a hole that contains the initial expression on which the completion is applied. We can model this using filled holes [28] and write $?_e$ for a hole containing the initial expression e. The $\mathsf{init}(?_e)$ operation then returns e alongside with an empty chain and the type of e.

As noted earlier, The Gamma does not embed query expressions directly into the generated code. It uses the same model as type providers and generates choices as members of types behind the scenes. We return to the differences between the two models in Section 5.

4.3 AI assistants

AI assistants guide the analyst through a data wrangling task. They generate a data cleaning script, taking into account constraints selected by the user. Most AI assistants obtain the script by performing statistical optimization with respect to a set of constraints specified by the user. That is, they look for an expression from the set of all possible expressions that optimizes some objective function that assigns score to the expression with respect to the given input data. Note that AI assistants do not iterate over all possible expressions. They use a machine learning method to approximate a solution to the problem.

An optimization-based AI assistant [36] thus provides another, very different, way of implementing the choose-your-own-adventure pattern. The assistant operates with respect to some input data X that does not change during the interaction and so we parameterize the choose-your-own-adventure calculus operations by the data. The input data X can be actual input data or a representative sample and so the AI assistant can be use past data to infer a cleaning script that will be used on new inputs.

The state σ consists of a set of constraints specified by the user. We write c for individual constraints and $\boldsymbol{c}$ for a set of constraints. The initial state is an empty set:

$$
\begin{aligned}
\sigma &= \{c_1, \ldots, c_n\} \\
\sigma_0 &= \emptyset
\end{aligned}
$$

Unlike in the previous examples, the crucial logic of an AI assistants is implemented in the choose operation. The operation runs the optimization algorithm to choose the best cleaning script for given constraints. Formally, this can be written using the arg max operator which finds an argument (an expression) for which the given function (scoring function) is maximized. The user-specified constraints can either restrict the set of possible expressions or influence the scoring function. More formally, we assume that:

- $E_{\boldsymbol{c}} \subseteq E$ is a set of expressions that respect constraints $\boldsymbol{c}$,
- $Q_{\boldsymbol{c}}(X, e)$ is a scoring function with respect to the constraints $\boldsymbol{c}$, which returns the score of an expression e, i.e., how good e is at cleaning the data X.

For a given set of constraints $\boldsymbol{c}$, the choose operation looks for $e \in E_{\boldsymbol{c}}$ with the largest score:

$$\mathsf{choose}_X(\boldsymbol{c}) = \arg\max_{e \in E_{\boldsymbol{c}}} Q_{\boldsymbol{c}}(X, e)$$

The actual implementation of the optimization uses various machine learning techniques to find the optimal expression. In case of datadiff, X is a pair of datasets X_1, X_2 to be reconciled. The AI assistant uses the Hungarian algorithm [43] to construct a matching of columns from X_1 and X_2. The generated expression is a sequence of patches that can be applied to X_2 in order to reconcile its structure with the sturcture of X_1. The constraints specified by the user restrict the space of possible column matchings and so they affect $E_{\boldsymbol{c}}$. The scoring function $Q_{\boldsymbol{c}}$ is independent of the constraints and computes a sum of distances between the statistical distributions of the columns from X_1 and a patched version of X_2.

The choices operation is responsible for generating possible constraints that the user may want to add to guide the inference. AI assistants typically offer the user options to prevent or adapt some aspect of the cleaning logic inferred by the system. For example, if datadiff matches two columns, it will offer a constraint to prevent the matching. It also generates constraints that let the user force a specific matching.

To implement $\mathsf{choices}_X$, optimization-based AI assistants first call $\mathsf{choose}_X(\sigma)$ to get the best expression e. Based on this, they generate possible constraints $c_1, c_2, \ldots$ that the user may want to choose from. The identifiers $\iota_1, \iota_2, \ldots$ provide a human-readable description of the constraints. Note that this operation is specific to the particular AI assistant. The $\mathsf{choices}_X$ operation then offers a list of constraint sets where the additional constraint is added to the previously collected set:

$$\mathsf{choices}_X(\boldsymbol{c}) = \{\iota_1 \mapsto \boldsymbol{c} \cup \{c_1\}, \iota_2 \mapsto \boldsymbol{c} \cup \{c_2\}, \ldots\}$$

The integration of an AI assistant, as described here, has to follow the external mode of embedding. The interaction with the assistant results in a cleaning script (expression), but there is no way of reconstructing the constrains used to guide the optimization. To support internal embedding, the choose operation would need to explicitly include the constraints in the resulting expression. However, rerunning the choose operation with the same constraints may result in a different cleaning script if the machine learning algorithm is probabilistic.

$$\boxed{\Gamma \vdash \tau \Rightarrow e}$$

$$\frac{}{\Gamma, x:\tau \vdash \tau \Rightarrow x}\ (\textsf{syn-var}) \qquad \frac{}{\Gamma \vdash \tau \Rightarrow ?_\tau}\ (\textsf{syn-hole})$$

$$\frac{\Gamma, x:\tau_1 \vdash \tau_2 \Rightarrow e}{\Gamma \vdash \tau_1 \to \tau_2 \Rightarrow \lambda x.e}\ (\textsf{syn-lambda}) \qquad \frac{\Gamma \vdash \tau_1 \to \tau_2 \Rightarrow e_1 \quad \Gamma \vdash \tau_1 \Rightarrow e_2}{\Gamma \vdash \tau_2 \Rightarrow e_1\ e_2}\ (\textsf{syn-app})$$

$$\frac{b \in \{\textsf{true}, \textsf{false}\}}{\Gamma \vdash \textsf{bool} \Rightarrow b}\ (\textsf{syn-bool}) \qquad \frac{\Gamma \vdash \tau \Rightarrow e_1 \quad \Gamma \vdash \tau \Rightarrow e_2 \quad \Gamma \vdash \textsf{bool} \Rightarrow e}{\Gamma \vdash \tau \Rightarrow \textsf{if}\ e\ \textsf{then}\ e_1\ \textsf{else}\ e_2}\ (\textsf{syn-cond})$$

Figure 5 Illustrative set of simple type-directed program synthesis rules

4.4 Theorem proving

In the previous three sections, we showed how existing formally well-documented systems fit the choose-your-own-adventure interaction pattern. Although interactive theorem provers and editors for dependently typed languages implement similar kinds of interactions, there is no well-documented system that exactly fits the pattern. The closest example is perhaps the recently envisioned mixed-mode interaction theorem prover [46]. Rather than reframing the implementation of an existing system, this section thus outlines a possible implementation.

Theorem provers. There are two approaches to interacting with an interactive theorem prover. In Coq, the user writes a sequence of tactics that transform proof goals. In Agda or Idris, the user writes a term, or program, of a type that represents the theorem. Interactive editors exist for both types of systems. For Coq, the `Company-Coq` [37, 4] extension offers auto-completion, which recommends available tactics, hypotheses and local definitions, but it does not filter them based on what is valid in a given context. For Idris, the interactive editor [7] offers a range of commands that transform the selected term by adding a case split, a missing case or by automatically searching for a proof. In Idris, the system produces valid completions, but those cover only a small number of situations.

Despite the different ways of working, an implementation of the choose-your-own-adventure pattern for both types of systems would be similar. Based on the sub-goal that the user is currently proving, the system would recommend a range of tactics that can be applied to the sub-goal. In the case of Coq, the selected tactic would be added to the sequence. In Idris, the selected tactic would be applied to transform the current term. The difference is in the mode of embedding. A system for Coq would provide internal embedding in that the selected option is added to the proof source code. (Much like selecting a completion when using type providers appends a member access.) A system for Idris-like language would use the tactic to transform the term, making it impossible to reconstruct the sequence of applied tactics as in the external mode of embedding.

Type-directed synthesis. Thanks to the equivalence between programs and proofs, techniques akin to tactic-based proof construction have also emerged in work on type-directed program synthesis [18]. As illustrated in Figure 5, the synthesis process can be described as a set of rules of the form $\Gamma \vdash \tau \Rightarrow e$ that describe ways of synthesizing expressions e of a type τ. Existing implementations of the mechanism typically aim to automate program synthesis and use more precise type information, such as refinement types [38] and graded types [15], or include examples [29]. However, the same rules could be used to guide an interactive choose-your-own-adventure system. If the interaction was invoked to fill a typed hole $?_\tau$ in a context Γ, the system could collect multiple e such that $\Gamma \vdash \tau \Rightarrow e$ and offer a choice of such options. Note that the definition in Figure 5 synthesizes sub-expressions recursively, but a choose-your-own-adventure system may always fill those with a typed hole using (syn-hole).

Formal model. From the perspective of user interaction, a proof assistant where a user interactively constructs a term of a given type is very similar to an interactive tool for type-directed program synthesis. The key difference being that theorem provers like Idris and Agda use rich dependent type theories.

For example, consider a system akin to Idris where the user aims to construct a term e of type τ. The term may contain typed holes written as $?_\tau$ and a relation $\Gamma \vdash \tau \Rightarrow e$ provides ways of synthesizing terms of type τ. We again write $E[?_\tau]$ for a completion context containing a (typed) hole; we assume that the variables Γ available in the completion context of the hole can be obtained using $\mathsf{vars}(E[\_])$.

To model a choose-your-own-adventure interaction akin to Idris, the state of the system would be the term e itself, initially a typed hole. The choices operation synthesizes possible completions using $\Rightarrow$ (restricted, e.g., to only generate terms of a certain maximum size) and offers the resulting terms as possible completions. The identifiers ι could be based either on the tactic name (rule name) or show a preview of the resulting term. The choices operation suggests ways to fill a hole in the term:

$$\begin{array}{rcl} \mathsf{choices}(E[?_\tau]) & = & \{\ \iota \mapsto e \mid \forall e\ .\ \mathsf{vars}(E[\_]) \vdash \tau \Rightarrow e_i\} \\ \mathsf{choose}(e) & = & e \end{array}$$

The choices operation synthesizes possible terms of a type required by the hole. Since the state σ is a term, the choose operation simply returns it. The definition models the external embedding of the interaction, i.e., a system that behaves according to Idris. It constructs the term, but does not record the completion choices. That said, a system akin to Coq that constructs a sequence of tactics could be modelled too if the state was a sequence of tactics and choices appended the available tactics as options to the end of the current sequence.

5 Properties

The choose-your-own-adventure calculus lets us precisely compare how how different programming systems interact with the user. We saw this in Section 3, which defines internal and external mode of embedding to distinguish between systems where the interaction leaves a reconstructible trace in the constructed program and systems where it does not. In this section, we make precise two properties that were introduced informally in the context of data exploration in The Gamma [33].

The choose-your-own-adventure system for data exploration in The Gamma is *correct*, meaning that all programs that a user can construct using the system, by repeatedly choosing from the auto-completion list, are well-typed. The system is also *complete*, meaning that the user can use auto-completion to construct all possible programs. In other words, there are no well-typed programs that cannot be constructed interactively, by repeatedly choosing options from the offered list of choices.

Correctness. The notions of correctness and completeness can be defined for any choose-your-own-adventure systems with respect to some system-specific distinction between correct and incorrect expressions. We write $\mathcal{E} \subseteq E$ for the subset of correct expressions.

For systems based on statically-typed programming languages, a reasonable choice of $\mathcal{E}$ is a set of all well-typed expressions. For some systems, we may additionally want the set of correct expressions $\mathcal{E}$ to be hole-free, i.e., only programs that can run (or represent complete proofs) are correct. For systems where the completion is string-based, we may treat all syntactically-correct programs as correct.

▶ **Definition 4** (Correctness). *Assume that $\mathcal{E} \subseteq E$ is a subset of correct expressions. A choose-your-own-adventure system is correct with respect to $\mathcal{E}$ if and only if:*

- $\forall \sigma_1, .., \sigma_n$ *and* $\iota_i, .., \iota_n$ *such that* $\iota_i \mapsto \sigma_i \in \mathsf{choices}(\sigma_{i-1})$ *it is the case that* $\mathsf{choose}(\sigma_i) \in \mathcal{E}$.

The definition states that, if we make any sequence of choices that start from an initial state σ_0 and result in intermediate states $\sigma_1, \ldots, \sigma_n$ then the programs we could generate from any of the intermediate states are correct.

The property depends on what we choose as the subset of correct expressions $\mathcal{E}$. Trivially, all systems are correct with respect to $\mathcal{E} = E$. However, the systems discussed in Section 4 are, with one exception, correct with respect to a non-trivial set of correct expressions:

- For the data exploration system discussed in Section 4.1, we say that correct expressions are those where the parameters of all operations are fully-specified. That is, no operation requires further arguments. With respect to this definition, the system is correct. However, this is the case only because the choose operation drops the last operation if it is not fully-specified. If choose returned all operations, including the partially constructed (but not yet completed) operation, the system would not be correct.
- For type providers (Section 4.2), correct expressions are those that are well-typed. With respect to this definition, the system is correct because the choices operation offers available members based on the type information. This reasoning also applies to the type provider behind The Gamma. In The Gamma, the generated members collect operation parameters and only invoke the operation once all parameters are known.
- In the case of AI assistants (Section 4.3) the correctness of the system depends on the expressions returned by the optimization algorithm (*arg max*) from the set of all possible cleaning scripts $E_{\boldsymbol{c}}$. In general, the algorithm can return any $e \in E_{\boldsymbol{c}}$ and so system

correctness is a matter of definition. The system is correct if and only if $E_{\boldsymbol{c}} \subseteq \mathcal{E}$ for all possible sets of constraints $\boldsymbol{c}$. In practice, it is more important that the constraints generated by $\mathsf{choices}$ are well-formed.

- For the interactive system based on type-directed synthesis (Section 4.4), correct expressions are those that are well-typed. The system is correct if the synthesis rules are sound [29], that is if $\Gamma \vdash \tau \Rightarrow e$ then also $\Gamma \vdash e : \tau$. Note that a correct choose-your-own-adventure system can be defined even using unsound synthesis rules – it would be sufficient to filter the recommended expressions in $\mathsf{choices}$ to the ones that are well-typed.

There may be useeful systems that violate the correctness property. An tool based on a large language model (LLM) may generate code with errors that the programmer can later correct. A more interesting case would be a data exploration system, like the one discussed above, where programs only become correct after multiple subsequent choices are made, for example to fully specify arguments of an operation.

Eventual correctness. The data exploration system discussed in Section 4.1 ensures correctness by dropping the last non-fully-specified operation in the choose operation. As a result, it is correct, but it does not support internal embedding. If the operation is dropped, we cannot reconstruct the full sequence of interactions from the generated source code. Generating code that includes the non-fully-specified operation allows external embedding, but makes the system incorrect. It would still satisfy a weaker definition of (eventual) correctness.

▶ **Definition 5** (Eventual correctness). *Assume that $\mathcal{E} \subseteq E$ is a subset of correct expressions, a choose-your-own-adventure system is eventually correct with respect to $\mathcal{E}$ if:*

- *For any sequence $\sigma_1, \ldots, \sigma_k$ and $\iota_1, \ldots, \iota_k$ such that $\forall i \in 1 \ldots k \;.\; \iota_i \mapsto \sigma_i \in \mathsf{choices}(\sigma_{i-1})$ there exists an extension $\sigma_{k+1}, \ldots, \sigma_n$ and $\iota_{k+1}, \ldots, \iota_n$ such that $\mathsf{choose}(\sigma_n) \in \mathcal{E}$ and $\forall i \in k{+}1 \ldots n \;.\; \iota_i \mapsto \sigma_i \in \mathsf{choices}(\sigma_{i-1})$.*

Eventual correctness models systems where some sequences of choices result in invalid programs, but it is always possible to make further choices to reach a valid program. It is always possible to turn an eventually correct system into a correct one:

1. As in the case of the data exploration, the system can remember the last state for which the choose operation returned a correct program and use it in choose until the next correct state is reached. This makes any eventually correct choose-your-own-adventure system correct, but it does not support internal embedding.
2. Alternatively, we can construct a system that collapses all sequence of temporarilly invalid states $\sigma_1, \ldots, \sigma_n$ identified by $\iota_1, \ldots, \sigma_n$ where $\forall i \in 1 \ldots n{-}1 \;.\; \mathsf{choose}(\sigma_i) \notin \mathcal{E}$ and $\mathsf{choose}(\sigma_n) \in \mathcal{E}$ into a single option $\iota' \mapsto \sigma_n$ where ι' is produced by concatenating identifiers $\iota_1, \ldots, \iota_n$. This makes the system correct and also preserves external embedding, but it potentially generates too many choices that are difficult to navigate.

There is more to be said about correctness of interactive programming systems, but the conceptual framework provided by the choose-your-own-adventure calculus makes it possible to take the first step. Similarly, the model lets us formally define completeness.

Completeness. The programming language used in The Gamma allows users to write scripts that also use let bindings and method calls. However, for chains of member accesses which the user can construct using a type provider, it is possible to construct any chain just by repeatedly choosing options from the offered list. This is captured as the completeness property of a choose-your-own-adventure system.

▶ **Definition 6** (Completeness). *Assume that $\mathcal{E} \subseteq E$ is a subset of correct expressions. A choose-your-own-adventure system is complete with respect to $\mathcal{E}$ if and only if:*

- $\forall e \in \mathcal{E} \,.\, \exists \sigma_1, .., \sigma_n$ *and* $\iota_i, .., \iota_n$ *such that* $\iota_i \mapsto \sigma_i \in \mathsf{choices}(\sigma_{i-1})$ *and* $e = \mathsf{choose}(\sigma_n)$.

A system is complete if, for any correct program, there is a sequence of choices that can be used to construct the given program. This is a more subtle property than correctness and it does not hold for all the examples dicusssed in Section 4.

- The data exploration system described in Section 4.1 would be complete only if the underlying query language had a fixed set of operations, a fixed set of aggregation operations (rather than letting users write their own) and a fixed set of values for each parameter (sorting by a key, but not based on a custom expression). A completion system for a more general-purpose query language would be incomplete.
- For type providers (Section 4.2), the completion mechanism is complete, because it offers all available members of the type (and "." has to be followed by a member access). Consequently, the type provider in The Gamma is also complete. Even if the underlying query language is more expressive, it is hidden from the user and the system offers all available members.
- For AI assistants (Section 4.3), the system offers a set of constraints based on the current inferred program. It does not let the user construct arbitrary constraints. Moreover, because the *arg max* operation used in choose performs statistical optimization, there is no guarantee that it can be used to generate a specific program. The system is only complete if it is possible to choose a constraint set $\boldsymbol{c}$ that restrict the set of programs $E_{\boldsymbol{c}}$ to a single given program. This is the case only for some AI assistants (e.g., datadiff, which always offers constraints to map column to any chosen other column).
- A type-directed synthesis system (Section 4.4), or an interactive theorem prover could offer all possible ways of filling a hole with an expression containing further holes as sub-expressions. This would make the system complete (up to renaming of variables this may introduce), but the great number of generated options would be impractical. A realistic system would only generate a subset of most useful proof/program steps and let the user write other steps manually (interactive proof construction in Idris can be seen as operating in this way).

Correctness and completeness are arguably both desirable properties of a choose-your-own-adventure system. Unlike for example type soundess, they are not strictly necessary in practice and are best seen as design trade-offs that designers should consider.

6 Applications

The choose-your-own-adventure calculus lets us treat a wide range of interactive programming systems as instances of the same general pattern. This makes it possible to discuss properties of the systems (as done in the previous section), reuse components in system implementations, but also transfer ideas across different domains. In this section, we discuss three ideas that emerged in the context of a specific interactive system, but could be applied to any system based on the choose-your-own-adventure pattern.

theorem sum-z-rh : **forall** n **exists** $n + (\mathsf{z}) = n$.
theorem sum-s-rh : **forall** $d_1 : n_1 + n_2 = n_3$ **exists** $n_1 + (\mathsf{s}\ n_2) = (\mathsf{s}\ n_3)$.

theorem sum-commutes : **forall** $d_1 : n_1 + n_2 = n_3$ **exists** $n_2 + n_1 = n_3$
  $d_2 : n_2 + n_1 = n_3$ **by induction on** d_1 :
  case rule

$$\frac{}{dzc : (\mathsf{z}) + n = n}\ \textsf{sum-z}$$

  is
    $dz_1 : n + (\mathsf{z}) = n$ **by theorem** sum-z-rh **on** n
  end case
  case rule

$$\frac{dsp : n_1' + n_2 = n_3'}{dsc : (\mathsf{s}\ n_1') + n_2 = (\mathsf{s}\ n_3')}\ \textsf{sum-s}$$

  is
    $ds_1 : n_2 + n_1' = n_3'$ **by induction hypothesis on** dsp
    $ds_2 : n_2 + (\mathsf{s} n_1') = (\mathsf{s} n_3')$ **by theorem** sum-s-rh **on** ds_1
  end case
  end induction
end theorem

Figure 6 A proof of commutativity of + in Peano arithmetic constructed using mixed-initiative interaction. Parts generated by the system are highlighted with gray background. After the user writes a theorem and specifies induction, the system completes the first case. User then specifies how to apply the induction hypothesis and the system completes the proof.

6.1 Mixed-initiative interaction

When using an interactive theorem prover, one typically constructs the proof manually until an automated strategy can fill in the remaining gaps. This is the case with Idris proof search, as well as Coq `auto` tactics. Richer ways of interacting exist [20], but are less common. We developed a prototype mixed-initiative theorem prover [report citation omitted] that supports a more collaborative way of working where the system completes some steps automatically, but defers back to the user when it gets stuck.

Mixed-initiative theorem proving. As an illustration of the mixed-initiative proving, consider Figure 6 which shows a proof of commutativity of + in Peano arithmetic. The figure shows a version of our example [report citation omitted],[3] simplified for brevity and written using the SASyLF notation [2]. After writing proofs of sum-z-rh and sum-s-rh (not shown), the user states sum-commutes and specifies the structure of the induction. They then invoke the automatic search, which completes the first case, but gets stuck in the second case, because it fails to apply the induction hypothesis (in our full example, the failure is more subtle). The user then specifies how to apply the induction hypothesis and the system automatically completes the proof.

The interaction can be revisited from the perspective of the choose-your-own-adventure

[3] Adapted from `https://github.com/boyland/sasylf/blob/master/examples/sum.slf`

system discussed in Section 4.4. An interactive theorem prover generates possible completions using available tactics and offers them to the user, who chooses a tactic and applies it to transform the proof. In the automatic mode, the interactive theorem prover recursively searches through the available choices. If it finds one that results in a complete proof, it stops. Otherwise, it completes a number of steps (determined by some heuristic) and defers back to the user who chooses the next step and completes the proof manually or invokes the automatic search again.

Generalised mixed-initiative interaction. The mixed-initiative mode of interaction combines manual interaction with automatic search. In order to support automatic search, the system needs a metric that determines whether a constructed program is correct (as when proving a given theorem) or whether it is an improvement over another equivalent program (for example when refactoring). This general way of working can be used in other interactive programming systems:

- A system for type-directed program synthesis could automatically synthesize parts of a program (e.g., pattern matching and implementation for some of the cases), but ask user for input in order to complete branches where solution was not found automatically.
- In data exploration, a system could perform automatic search through the available operations in order to transform data into a more regular format, according to a suitability metric as done, for example, in the Proactive Wrangling system [14].

We can understand how mixed-initiative interactive systems as extensions built on top of the choose-your-own-adventure calculus. A mixed-initiative interactive system defines an operation suggest that recommends a sequence of choices for a given state.

▶ **Definition 7** (Mixed-initiative system)**.** *A choose-your-own-adventure system supports mixed-initiative mode of interaction if it is equipped with an operation* suggest *such that:*
- $\mathsf{suggest}(\sigma_0) = \iota_1, \ldots, \iota_n$ *such that* $\forall i \in 1 \ldots n \;.\; \iota_i \mapsto \sigma_i \in \mathsf{choices}(\sigma_{i-1})$.

The definition only requires that the suggested sequence of choices is valid, but it does not specify how exactly the system should make the recommendations. This is specific to a particular system. An interactive theorem prover will try to find a proof or solve sub-goals, whereas data wrangling system may try to improve the structure of data. In practice, the suggestion should improve the quality of the constructed program so that $\mathsf{choose}(\sigma_n)$ is better than $\mathsf{choose}(\sigma_0)$, but we leave the definition flexible to accommodate a broader range of potential uses.

```
You are helping user to complete a task in an interactive programming environment.
The user's query is: "Give me the athlete with the largest number of gold medals."

The query built so far is: "olympics"."group data"."by Athlete".

The environment offers the user possible options. Choose an option that the should
be applied to the current dataset:

1.  count distinct Athlete
2.  count distinct Discipline
    [multiple options omitted]
13. count all
14. concatenate Athlete
15. concatenate Discipline
    [multiple options omitted]
23. sum Gold
24. sum Silver
25. sum Year

You should answer with the number of the option and no further explanation.
```

Figure 7 A prompt to complete a data exploration query based on a natural language question, which uses LLM to choose from the available completion options.

6.2 Language model-based completion

Large language models can assist with data exploration by generating snippets of code based on natural language description of the problem [47]. A recognized drawback of this approach is that the user may gain only cursory understanding of the code and overlook errors. Various systems address this by generating explanations alongside with code [27]. We developed a system that provides LLM-based natural language assistance for The Gamma [report citation omitted] based on a mechanism that assists the user and is also applicable to other choose-your-own-adventure systems.

LLM assistant for The Gamma. Our system [reprot citation omitted] integrates a type provider like the one used in The Gamma with a large language model (GPT-3.5 and GPT-4). It lets user ask a natural language query and then recommends choices that the user should make in order to answer the query. Our system does not use LLM to generate code, but to recommend an option offered by the type provider. To do so, we construct a prompt as shown in Figure 7, which asks the LLM to recommend the next choice. In this example, the LLM is able to reliably respond with "23", which is the correct choice. Our system then pre-selects it in the completion list offered to the user.

In contrast to one-shot generation of Python or SQL code snippets, our approach guides the user through the program construction, providing them with an opportunity to review and check if the program logic matches their expectations.

In addition to the basic prompting strategy outlined above, we evaluated a strategy where the LLM is provided with lookahead information, i.e., for each of the choices, we also include a list of the choices that will available subsequently (formatted as either inline information in parentheses or as a nested bullet-point list). We use two different data sources to evaluate the system, each with 10 queries for which we manually determine the correct choices. To compare the results, we compute a score for each query by following the correct path, asking for an LLM recommendation in each step and computing the ratio of correct choices. Figure 8 shows the average scores for five different strategies, using two different LLM engines.

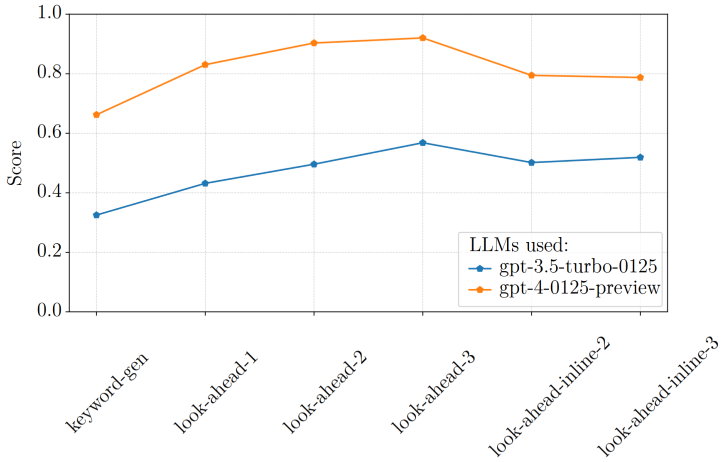

Figure 8 A plot showing the average scores for two different large language models (GPT-3.5 and GPT-4) solving 20 challenge problems using five different prompting strategies.

The results suggest that, when using a more advanced LLM system, the ratio of correct recommendations is high-enough to be practically useful. It also shows that providing more information to the LLM through the lookahead mechanism improves the quality of the recommendation, in a consistent way across multiple LLM systems. Arguably, offering a recommendation is preferrable over providing an opaque solution, because it keeps the human in the loop [40] and avoids "ironies of automation" [5] including deskilling.

Generalised LLM assistant. The LLM-based recommendation engine developed for The Gamma can be implemented for any choose-your-own-adventure system. As can be seen in Figure 7, the information needed to construct an LLM prompt comprise only the natural language query entered by the user, a sequence of previously made choices $\iota_1, \ldots, \iota_{n-1}$ and the identifiers $\iota_n, \iota'_n, \iota''_n, \ldots$ of the choices offered by the choices operation. As the LLM-based recommendations are based on natural language analysis of the prompt, the quality of the recommendations depends on how semantically meaningful the generated identifiers are.

An LLM-based recommendation engine can be useful for a number of the choose-your-own-adventure interactive systems discussed in this paper:

- We demonstrated the usefulness of the system for data exploration in The Gamma. Type providers for structured data [35] typically generate small number of choices that the user can navigate without assistance, but navigating the schemas generated by type providers for semantic knowledge bases [44] may be simplified through a natural language query.
- The datadiff AI assistant discussed in Section 2 matches columns based on statistical distribution and ignores column names. An LLM-based recommendation engine is useful in this case, because it is able to recommend the option Don't match 'Urban.rural' and 'Nation' based solely on its name. For other AI assistants, the LLM-based recommendation engine would need access to the input data in order to be useful.
- The problem of predicting steps in order to construct a proof is known as proofstep generation in literature focused on deep learning for theorem proving [19]. Although most systems use models trained specifically for the task, there is also interest in using general-purpose LLMs [48]. Our approach suggests a potential prompting strategy.

Figure 9 Programming by demonstration in Histogram – (1) the user constructs program to load data, (2) then they filter data using graphical interface, which (3) records an operation corresponding to the interaction in code.

It is interesting to note that the choose-your-own-adventure calculus provides a common structure that makes it possible to integrate multiple different approaches to assisting human in program construction, ranging from automatic search (if we can provide a suitable heuristic) to the use of LLMs (if generated choices carry enough semantic information). It can also serve as the basis for visual programming environments discussed next.

6.3 Direct manipulation

The Histogram programming environment [31] provides a code completion mechanism that can be used both to choose operations (as with type providers) and to specify their parameters (by offering available values in context as possible arguments). It makes it possible to evaluate sub-expressions at a fine-grained level and also refines type information based on runtime values. However, it also supports programming by demonstration [10, 17].

As illustrated in Figure 9, when the user loads data, they can manipulate it in a table through a direct manipulation interface [41]. In Histogram, interacting with the table triggers a sequence of operations that construct code. A similar functionality is available in The Gamma, where operations for manipulating tabular data can be selected not only through auto-completion, but also by interacting with the live preview.

The choose-your-own-adventure calculus enables a more general perspective on the programming by demonstration interaction as implemented by Histogram. When interacting with a live preview, the user is using a graphical interface to choose an option offered by choices. The operation appends an item to the sequence of choices, updates the system state and generates a new preview where the choice is reflected (e.g., by filtering the table).

An interesting question is whether a program can be constructed solely through direct manipulation, by interacting with the live previews. For this to be possible, the live preview needs to offer links to all the possible options offered by choices. This requires a careful design of live previews. In Histogram (Figure 9), the live preview offers interactive elements for filtering based on equality test (row of drop-downs), sorting (up/down arrows) and indexing (indices in the "#" column), but some operations (such as aggregation) can only be constructed through code. Previews in The Gamma provide interactive elements for all options, but those cannot always be integrated with the data display. For example, after choosing a grouping key (Figure 10), the preview shows the key column, but other columns are hidden. Specifying an aggregation thus requires an additional (somewhat arbitrary) "+" button that provides access to the available choices.

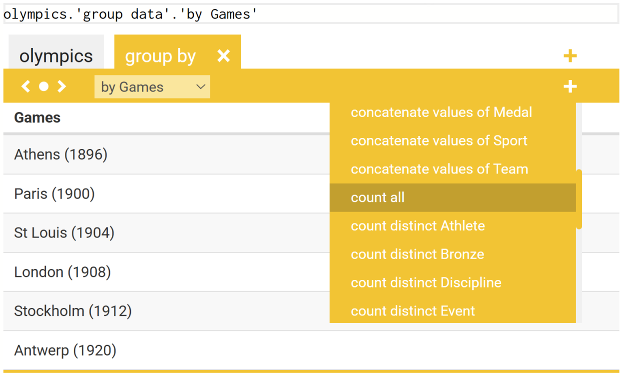

Figure 10 Specifying aggregation in The Gamma through a live preview. Grouping keys can be selected from a drop-down, while aggregations have to be added through a menu.

We can use the choose-your-own-adventure calculus to make the question whether all programs can be constructed through direct manipulation more precise. Assume $\mathsf{preview}(e) = p$ is a preview constructed for an expression e and $\mathsf{links}(p) = \iota_1, \ldots, \iota_n$ are identifiers that can be invoked through the preview (akin to links in a hypertext document).

▶ **Definition 8** (Direct manipulation). *A choose-your-own-adventure system with previews defined by* preview *and* links *supports full direct manipulation if:*

- $\forall \sigma, e$ *such that* $\mathsf{choose}(\sigma) = e$ *it is the case that* $\forall \iota' \mapsto \sigma' \in \mathsf{choices}(\sigma)\,.\,\iota' \in \mathsf{links}(\mathsf{preview}(e))$.

A system supports full direct manipulation if any program can be constructed by interacting with the live previews, created based on the gradually constructed programs. As illustrated by The Gamma, this can always be achieved by listing the options in a menu, but it is more interesting to see if the links can be directly embedded in the preview as in the data table interface of Histogram.

The idea of directly mapping elements in a user interface to underlying structure of the programming system has previously been implemented in pioneering user interface systems including the Alternate Reality Kit [42] and the Morphic framework for Self and Squeak [23, 22]. In those systems, user interface interactions directly map to messages sent to the underlying object and the halo element in Morphic plays a similar role to the additional menu in The Gamma. It remains to be seen if the choose-your-own-adventure calculus can provide a new way of looking at those systems.

7 Limitations

TODO: Is this too general to be practically useful? E.g. LLM for interactive theorem provers arguably needs more specific information.

TODO: What if the list of choices is infinite?

TODO: How do you go back and revisit earlier choices? Backtracking in theorem provers or fixing earlier part of the query in data exploration. Sometimes, you can retrace the steps, but they may change...

Related work - mbark for emacs, VS code command palette

Alphageometry paper - follows a similar way of working to choose-your-own-adventure

8 Conclusions

TODO: Some conclusions 1) lets us ask all those questions 2) let ideas cross between domains

References

1 Michael D. Adams, Eric Griffis, Thomas J. Porter, Sundara Vishnu Satish, Eric Zhao, and Cyrus Omar. Grove: A bidirectionally typed collaborative structure editor calculus. *Proc. ACM Program. Lang.*, 9(POPL), January 2025. doi:10.1145/3704909.

2 Jonathan Aldrich, Robert J. Simmons, and Key Shin. Sasylf: an educational proof assistant for language theory. In *Proceedings of the 2008 International Workshop on Functional and Declarative Programming in Education*, FDPE '08, page 31–40, New York, NY, USA, 2008. Association for Computing Machinery. doi:10.1145/1411260.1411266.

3 Thorsten Altenkirch, Veronica Gaspes, Bengt Nordström, and Björn von Sydow. A user's guide to alf. Technical report, Chalmers University of Technology, Sweden, 1994. Unpublished Draft. URL: https://people.cs.nott.ac.uk/psztxa/publ/alf94.pdf.

4 David Aspinall. Proof general: A generic tool for proof development. In Susanne Graf and Michael Schwartzbach, editors, *Tools and Algorithms for the Construction and Analysis of Systems*, pages 38–43, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.

5 Lisanne Bainbridge. Ironies of automation. In G. Johannsen and J.E. Rijnsdorp, editors, *Analysis, Design and Evaluation of Man–Machine Systems*, pages 129–135. Pergamon, 1983. doi:10.1016/B978-0-08-029348-6.50026-9.

6 Alan F. Blackwell and Thomas R. G. Green. Notational systems – the cognitive dimensions of notations framework. In John M. Carroll, editor, *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*, pages 103–134. Morgan Kaufmann, San Francisco, 2003.

7 Edwin Brady. *The Idris Programming Language*, pages 115–186. Springer International Publishing, Cham, 2015. doi:10.1007/978-3-319-15940-9_4.

8 Edwin Brady. Idris 2: Quantitative type theory in practice. In Anders Møller and Manu Sridharan, editors, *35th European Conference on Object-Oriented Programming (ECOOP 2021)*, volume 194 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 9:1–9:26, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ECOOP.2021.9.

9 E. F. Codd. A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377–387, June 1970. doi:10.1145/362384.362685.

10 A. Cypher and D.C. Halbert. *Watch what I Do: Programming by Demonstration.* MIT Press, 1993. URL: https://books.google.cz/books?id=Ggzjo0-W1y0C.

11 Damian Frölich and L. Thomas van Binsbergen. On the soundness of auto-completion services for dynamically typed languages. In *Proceedings of the 23rd ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences*, GPCE '24, page 107–120, NY, USA, 2024. Association for Computing Machinery. doi:10.1145/3689484.3690734.

12 Richard P. Gabriel. The structure of a programming language revolution. In *Proceedings of the ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software*, Onward! 2012, page 195–214, New York, NY, USA, 2012. Association for Computing Machinery. doi:10.1145/2384592.2384611.

13 Thomas R. G. Green. Cognitive dimensions of notations. In A. Sutcliffe and L. Macaulay, editors, *People and Computers V: Proceedings of the Fifth Conference of the British Computer Society*, pages 443–460, Cambridge, UK, 1989. Cambridge University Press.

14 Philip J. Guo, Sean Kandel, Joseph M. Hellerstein, and Jeffrey Heer. Proactive wrangling: mixed-initiative end-user programming of data transformation scripts. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, page 65–74, New York, NY, USA, 2011. Association for Computing Machinery. doi:10.1145/2047196.2047205.

15 Jack Hughes and Dominic Orchard. Program synthesis from graded types. In Stephanie Weirich, editor, *Programming Languages and Systems*, pages 83–112, Cham, 2024. Springer Nature Switzerland.

16 Joel Jakubovic, J. Edwards, and T. Petricek. Technical dimensions of programming systems. *Art Sci. Eng. Program.*, 7(3), 2023. doi:10.22152/PROGRAMMING-JOURNAL.ORG/2023/7/13.

17 Sean Kandel, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer. Wrangler: interactive visual specification of data transformation scripts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 3363–3372, New York, NY, USA, 2011. Association for Computing Machinery. doi:10.1145/1978942.1979444.

18 Tristan Knoth. *Type-Directed Program Synthesis*. PhD thesis, University of California, San Diego, 2023. URL: https://escholarship.org/uc/item/4g11m7rq.

19 Zhaoyu Li, Jialiang Sun, Logan Murphy, Qidong Su, Zenan Li, Xian Zhang, Kaiyu Yang, and Xujie Si. A survey on deep learning for theorem proving. In *First Conference on Language Modeling*, 2024. URL: https://openreview.net/forum?id=zlw6AHwukB.

20 Helen Lowe and David Duncan. Xbarnacle: Making theorem provers more accessible. In William McCune, editor, *Automated Deduction—CADE-14*, pages 404–407, Berlin, Heidelberg, 1997. Springer Berlin Heidelberg.

21 Lena Magnusson and Bengt Nordström. The alf proof editor and its proof engine. In Henk Barendregt and Tobias Nipkow, editors, *Types for Proofs and Programs*, pages 213–237, Berlin, Heidelberg, 1994. Springer Berlin Heidelberg.

22 John Maloney. An introduction to morphic: The squeak user interface framework. In Mark Guzdial and Kim Rose, editors, *Squeak: Open Personal Computing and Multimedia*. Prentice Hall, 2001.

23 John H. Maloney and Randall B. Smith. Directness and liveness in the morphic user interface construction environment. In *Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology*, UIST '95, page 21–28, New York, NY, USA, 1995. Association for Computing Machinery. doi:10.1145/215585.215636.

24 Mikaël Mayer, Viktor Kuncak, and Ravi Chugh. Bidirectional evaluation with direct manipulation. *Proc. ACM Program. Lang.*, 2(OOPSLA), October 2018. doi:10.1145/3276497.

25 Conor McBride. *Dependently Typed Functional Programs and their Proofs*. PhD thesis, University of Edinburgh, 1999. URL: https://era.ed.ac.uk/handle/1842/374.

26 Jaroslav Nešetřil. Strom jako matematická struktura – i v umění. KAM Series 742, Department of Applied Mathematics, Charles University, 2005. URL: https://kam.mff.cuni.cz/~kamserie/serie/clanky/2005/s742.pdf.

27 Farhad Nooralahzadeh, Yi Zhang, Jonathan Furst, and Kurt Stockinger. Explainable multimodal data exploration in natural language via llm agent, 2024. URL: https://arxiv.org/abs/2412.18428, arXiv:2412.18428.

28 Cyrus Omar, Ian Voysey, Ravi Chugh, and Matthew A. Hammer. Live functional programming with typed holes. *Proc. ACM Program. Lang.*, 3(POPL), January 2019. doi:10.1145/3290327.

29 Peter-Michael Osera and Steve Zdancewic. Type-and-example-directed program synthesis. In *Proceedings of the 36th ACM SIGPLAN Conference on Programming Language Design and Implementation*, PLDI '15, page 619–630, New York, NY, USA, 2015. Association for Computing Machinery. doi:10.1145/2737924.2738007.

30 Tomas Petricek. Data exploration through dot-driven development. In Peter Müller, editor, *31st European Conference on Object-Oriented Programming, ECOOP 2017, June 19-23, 2017, Barcelona, Spain*, volume 74 of *LIPIcs*, pages 21:1–21:27. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017. doi:10.4230/LIPICS.ECOOP.2017.21.

31 Tomas Petricek. Histogram: You have to know the past to understand the present. In *Workshop on Live Programming (LIVE 2019)*, Athens, Greece, 2019. Presented at the Workshop on Live Programming, Published online. URL: https://tomasp.net/histogram.

32 Tomas Petricek. Foundations of a live data exploration environment. *Art Sci. Eng. Program.*, 4(3):8, 2020. doi:10.22152/PROGRAMMING-JOURNAL.ORG/2020/4/8.

33 Tomas Petricek. The gamma: Programmatic data exploration for non-programmers. In Paolo Bottoni, Gennaro Costagliola, Michelle Brachman, and Mark Minas, editors, *2022 IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC 2022, Rome, Italy, September 12-16, 2022*, pages 1–7. IEEE, 2022. doi:10.1109/VL/HCC53370.2022.9833134.

34 Tomas Petricek, James Geddes, and Charles Sutton. Wrattler: Reproducible, live and polyglot notebooks. In *10th USENIX Workshop on the Theory and Practice of Provenance (TaPP 2018)*, London, July 2018. USENIX Association. URL: `https://www.usenix.org/conference/tapp2018/presentation/petricek`.

35 Tomas Petricek, Gustavo Guerra, and Don Syme. Types from data: making structured data first-class citizens in F#. In *Proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation*, PLDI '16, page 477–490, New York, NY, USA, 2016. Association for Computing Machinery. `doi:10.1145/2908080.2908115`.

36 Tomas Petricek, Gerrit J. J. van den Burg, Alfredo Nazábal, Taha Ceritli, Ernesto Jiménez-Ruiz, and Christopher K. I. Williams. AI assistants: A framework for semi-automated data wrangling. *IEEE Trans. Knowl. Data Eng.*, 35(9):9295–9306, 2023. `doi:10.1109/TKDE.2022.3222538`.

37 Clément Pit-Claudel and Pierre Courtieu. Company-coq: Taking proof general one step closer to a real ide. In *Second International Workshop on Coq for PL (CoqPL '16)*, January 2016. URL: `https://dspace.mit.edu/handle/1721.1/101149.2`.

38 Nadia Polikarpova, Ivan Kuraj, and Armando Solar-Lezama. Program synthesis from polymorphic refinement types. In *Proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation*, PLDI '16, page 522–538, New York, NY, USA, 2016. Association for Computing Machinery. `doi:10.1145/2908080.2908093`.

39 Mark Seemann. *Code That Fits in Your Head: Heuristics for Software Engineering*. Addison-Wesley Professional, Boston, 2021.

40 Abigail Sellen and Eric Horvitz. The rise of the ai co-pilot: Lessons for design from aviation and beyond. *Commun. ACM*, 67(7):18–23, July 2024. `doi:10.1145/3637865`.

41 Ben Shneiderman. Direct manipulation: A step beyond programming languages. *Computer*, 16(8):57–69, 1983. `doi:10.1109/MC.1983.1654471`.

42 Randall B. Smith. Experiences with the alternate reality kit: an example of the tension between literalism and magic. In *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, CHI '87, page 61–67, New York, NY, USA, 1986. Association for Computing Machinery. `doi:10.1145/29933.30861`.

43 Charles A. Sutton, Timothy Hobson, James Geddes, and Rich Caruana. Data Diff: Interpretable, Executable Summaries of Changes in Distributions for Data Wrangling. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2279–2288. ACM, 2018. `doi:10.1145/3219819.3220057`.

44 Don Syme, Keith Battocchi, Kenji Takeda, Donna Malayeri, Jomo Fisher, Jack Hu, Tao Liu, Brian McNamara, Daniel Quirk, Matteo Taveggia, Wonseok Chae, Uladzimir Matsveyeu, and Tomas Petricek. F# 3.0 - strongly-typed language support for internet-scale information sources. Technical Report MSR-TR-2012-101, September 2012. URL: `https://www.microsoft.com/en-us/research/publication/f3-0-strongly-typed-language-support-for-internet-scale-information-sources/`.

45 Don Syme, Keith Battocchi, Kenji Takeda, Donna Malayeri, and Tomas Petricek. Themes in information-rich functional programming for internet-scale data sources. In Evelyne Viegas, Karin K. Breitman, and Judith Bishop, editors, *Proceedings of the 2013 Workshop on Data Driven Functional Programming, DDFP 2013, Rome, Italy, January 22, 2013*, pages 1–4. ACM, 2013. `doi:10.1145/2429376.2429378`.

46 Jan Liam Verter and Tomas Petricek. Don't call us, we'll call you: Towards mixed-initiative interactive proof assistants for programming language theory. *CoRR*, abs/2409.13872, 2024. Presented at the 5th International Workshop on Human Aspects of Types and Reasoning Assistants (HATRA 2024). `arXiv:2409.13872`, `doi:10.48550/arXiv.2409.13872`.

47 Pengcheng Yin, Wen-Ding Li, Kefan Xiao, Abhishek Rao, Yeming Wen, Kensen Shi, Joshua Howland, Paige Bailey, Michele Catasta, Henryk Michalewski, Oleksandr Polozov, and Charles Sutton. Natural language to code generation in interactive data science notebooks. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*,

pages 126–173, Toronto, Canada, July 2023. Association for Computational Linguistics. doi:10.18653/v1/2023.acl-long.9.

48 Shizhuo Dylan Zhang, Talia Ringer, and Emily First. Getting more out of large language models for proofs. *CoRR*, abs/2305.04369, 2023. Presented at the 8th Conference on Artificial Intelligence and Theorem Proving (AITP 2023). arXiv:2305.04369, doi:10.48550/arXiv.2305.04369.