Understanding How In-Visualization Provenance Can Support Trade-off Analysis

Mehdi Chakhchoukh, Nadia Boukhelifa, Anastasia Bezerianos

This is the author version of the article. The final version is published in IEEE Transactions on Visualization and Computer Graphics 2022, at: 10.1109/TVCG.2022.31710741

Understanding How In-Visualization Provenance Can Support Trade-off Analysis

Mehdi Chakhchoukh, Nadia Boukhelifa, and Anastasia Bezerianos

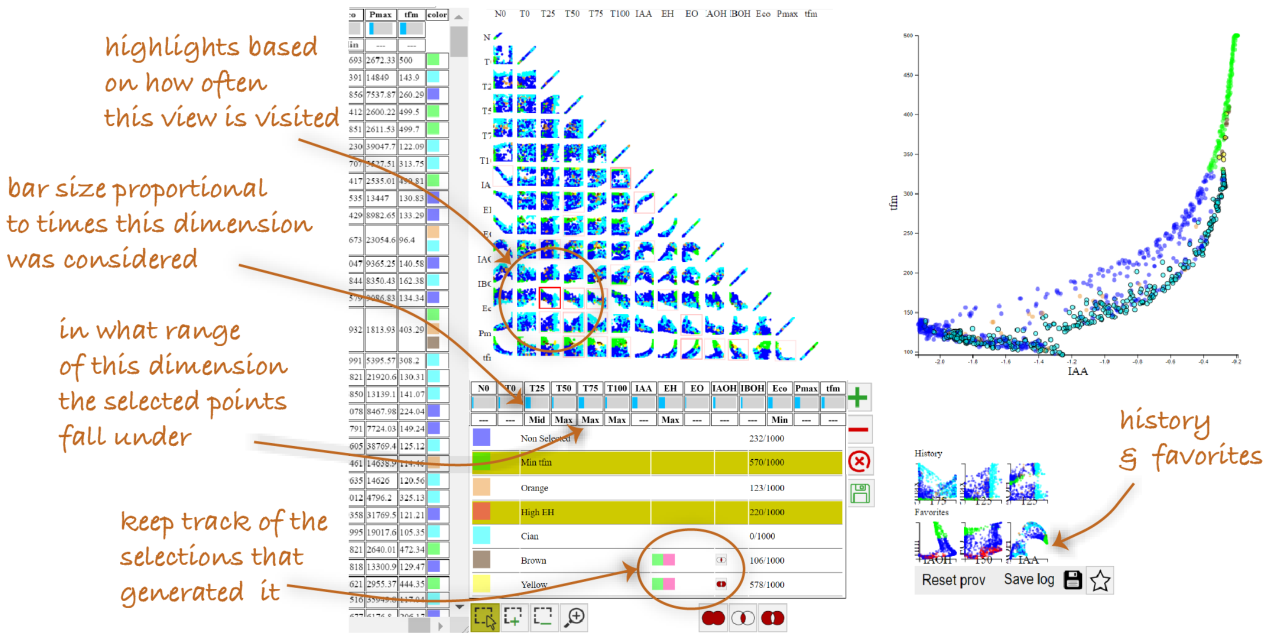


Fig. 1. The VisProm technology probe includes several in-visualization provenance views to aid trade-off analysis.

Abstract—In domains such as agronomy or manufacturing, experts need to consider trade-offs when making decisions that involve several, often competing, objectives. Such analysis is complex and may be conducted over long periods of time, making it hard to revisit. In this paper, we consider the use of analytic provenance mechanisms to aid experts recall and keep track of trade-off analysis. We implemented VisProm, a web-based trade-off analysis system, that incorporates in-visualization provenance views, designed to help experts keep track of trade-offs and their objectives. We used VisProm as a technology probe to understand user needs and explore the potential role of provenance in this context. Through observation sessions with three groups of experts analyzing their own data, we make the following contributions. We first, identify eight high-level tasks that experts engaged in during trade-off analysis, such as locating and characterizing interest zones in the trade-off space, and show how these tasks can be supported by provenance visualization. Second, we refine findings from previous work on provenance purposes such as recall and reproduce, by identifying specific objects of these purposes related to trade-off analysis, such as interest zones, and exploration structure (e.g., exploration of alternatives and branches). Third, we discuss insights on how the identified provenance objects and our designs support these trade-off analysis tasks, both when revisiting past analysis and while actively exploring. And finally, we identify new opportunities for provenance-driven trade-off analysis, for example related to monitoring the coverage of the trade-off space, and tracking alternative trade-off scenarios.

Index Terms—Provenance, visualization, trade-offs, multi-criteria, decision making, qualitative study.

1 Introduction

Domains such as agronomy and manufacturing, often use simulation models that represent entities and parameters of complex biological or mechanical processes, and their relationships. Experts explore the results of these models, reach insights and make decisions about how to optimize their processes. For example, an agronomic engineer who wants to propose sustainable, but robust, wheat fertilization strategies to farmers, needs to account for the wheat growth process and how it is affected by soil and weather conditions, as well as the impact of the chosen fertilization strategy on the environment. Thus experts have to deal with *trade-offs*, attempting to consider and reconcile multiple competing objectives in a single investigation [1]. For instance, the agronomic engineer needs to find fertilization strategies that on the one hand

- *M. Chakhchoukh is with Univ. Paris-Saclay, CNRS, INRIA, & UMR MIA-Paris, AgroParisTech, INRAE. E-mail: mehdi.chakhchoukh@universite-paris-saclay.fr.*
- *N. Boukhelifa is with UMR MIA-Paris, AgroParisTech, INRAE, Univ. Paris-Saclay. E-mail: nadia.boukhelifa@inrae.fr.*
- *A. Bezerianos is with Univ. Paris-Saclay, CNRS, INRIA. E-mail: anastasia.bezerianos@universite-paris-saclay.fr*

maximize yield, and on the other hand reduce the amount of supplied fertilizers, and nitrogen loss to the plant.

This type of trade-off analysis can be complex as it captures experts requirements and preferences across several objectives and often invovles multiple experts working together to understand different aspects of the data [2]. Depending on the domain, datasets can include tens of dimensions and hundreds or thousands of datapoints. Using their domain knowledge, experts analyze these large multi-dimensional spaces, prioritizing requirements and balancing subjective preferences across several competing objectives. This complex analysis is conducted over long periods of time and several exploration sessions [2], as multiple experts try to explore alternatives and reach common ground , making this type of analysis hard to resume and revisit. We explore the use of analytic provenance visualizations [3], [4], [5], [6], to aid experts keep track, resume and revisit their trade-off analysis sessions. How provenance can support trade-off analysis tasks is a question that has not been considered before, and poses unique challenges as analysts need to keep track of the objectives they have already considered, their importance, and the trade-offs they represent [7]. Inspired by past observations on trade-off analysis challenges [2], we designed and developed a web-based trade-off analysis tool (VisProm) that embeds provenance visualizations directly inside the analysis environment, i.e., *in-visualization provenance*. The tool was designed specifically with trade-off analysis in mind, and simplicity when it comes to the proposed provenance views. Since no previous study looked at the role of in-visualizaton provenance in supporting trade-off analysis, our hypothesis here is that if such simple views are helpful for trade-off analysis, then richer provenance views may be investigated in the future. We use the VisProm system as a technology probe [8]; a prototype system designed to be field tested in order to inspire users and designers to think about the benefits and limits of provenance support in trade-off analysis, rather than a fully-fledged trade-off analysis system. Our goal is to bring new insights on analytical provenance usage in real world analysis scenarios, both during and after data exploration, as well as to identify new opportunities for future trade-off provenance designs.

This paper describes findings from an observational study that investigates the use of provenance in trade-off analysis, where three groups of experts from different domains used VisProm to explore their own data. Our contributions are: **(i)** the delineation of eight trade-off analysis tasks extending previous work in operations research [7]; **(ii)** the specification of the general provenance purposes [4], through eight provenance objects, that our participants appear to focus on during trade-off analysis; **(iii)** insights on how the identified provenance objects and our implemented designs support these trade-off analysis tasks during a-posteriori exploration and active analysis; and **(iv)** the identification of specific needs and opportunities for the use of provenance during trade-off analysis, that refine or go beyond those identified in previous provenance work [4] [6].

2 Related Work

Since our goal is to study provenance in the context of trade-off analysis, we review research on provenance visualization for sensemaking (sections 2.1), and decision making and trade-off analysis (section 2.2). We summarize recent work from these two fields separately as, to the best of our knowledge, there is little work that addresses both at the same time.

2.1 Provenance

Xu et al. define provenance information as the history of the data and reasoning involved [during complex sensemaking tasks] and the context within which sensemaking was performed [9]. It is invaluable as a means to help analysts retrace their exploration and how it led to insights. As Ragan et al. explain [4], the types of provenance information that may be of interest to analysts can vary greatly. It can range from keeping track of low level information such as data transformations, view changes and user interactions; to high level such as capturing insights reached during the analysis, or the rational behind decisions and hypotheses.

Visual analysis tools often capture basic view and interaction events that analysts can access in the form of a history or action log. But several systems go beyond this basic provenance capture and visualization. Past work has explored how to revisit, operate on (search, filter, annotate) [10] and more generally how to visualize and curate past history [11]. It also studied ways to help connect exploration steps [12], and aid analysts summarize and hand-off their analysis to colleagues [13]. A recent survey on the topic in interaction provenance provides further reading [14].

More relevant to our work are tools that provide *in-situ* provenance visualizations (or in-visualization provenance), augmenting existing parts of the interface with simple visual marks to encode provenance information. The first such instance is the idea of "visit ware" [15], that applies a fish-eye lens on a graph to magnify the nodes that are more visited. Later BookVoyager [16] attempted on the contrary to de-emphasize the most visited timeseries in a line chart by graying them out, in order to encourage users to visit new parts of the visualization ("road-less-traveled navigation").

Scented Widgets [17] introduce a framework for enhancing interface components (such as buttons or sliders) with information that can aid navigation. While the concept is general, the work suggests several examples of visualizing provenance information, such as changing the color of a button based on the frequency of use, or keeping track of datasets visited in a drop-down list.

HindSight [18] encoded interaction history directly on visualizations (in-visualization history), for example by changing the opacity of the most visited chart in a small multiples list, or the color and width of the most visited line in a line chart. In an online study they observed that adding such interaction history increased user interaction with the interface and may have led users towards typically unexplored areas of the visualization.

Most recently Lumos [19], [20] added to scatterplots (among other views) in-situ visualizations of interaction traces: real-time ones, such as coloring individual data-points or attributes to encode frequency of interaction; and summative ones to show the distribution of data-points that users interacted with compared to a target distribution on a dimension. Their goal was to enhance awareness of biases [19] and mitigate them [20] during exploration and decision making. In a series of laboratory and crowdsource experiments, they found that these traces increased awareness of biases [19], and somewhat mitigated unintentional bias but enabled intentional bias [20].

The provenance designs introduced in our technology probe follow this *in-situ* approach, augmenting existing components of the interface with simple visual indicators of provenance information. Nevertheless, we contribute a study that uses in-situ visualizations to observe real analysts: our experts revisit their own past analysis and data, days or even months later; and also use our provenance visualizations during active exploration. Moreover, we introduce in-situ visualizations in an environment targeting a specific type of visual analysis: trade-off exploration.

Purposes of provenance

While provenance support has not been studied in the context of trade-off analysis, past work has highlighted what we should expect provenance to be used for. Beyond provenance information types, Ragan et al. [4] have also identified the following reasons why analysts use provenance (purposes): ***Recall:*** maintaining and recovering awareness and memory of the current and past states of the analysis; ***Replication:*** reproducing aspects of a previous analysis (steps or workflow); ***Action recovery:*** undoing/redoing past actions during the analysis; ***Collaborative communication:*** sharing data, information and ideas with colleagues; ***Presentation:*** summarizing and sharing the results or processes of the analysis with someone not involved in the analysis; and ***Meta-analysis:*** reviewing the analysis process itself. These general provenance purposes serve as the starting point for our own analysis (section 5.2). In our work we contribute a refinement of these purposes based on specific trade-off analysis needs.

2.2 Decision Making & Trade-off Analysis

In domains such as agronomy or engineering design, experts often explore the results of large multi-dimensional search spaces derived from model simulations. Their goal is to identify solutions (or simulations) that fit a set of criteria they have, that can be competing - representing *trade-offs*. This goal is related to multi-attribute choice tasks that focus on finding the best option out of a set of alternatives defined across multiple attributes [21], [22]. Nevertheless, trade-off analysis does not consist of finding one single best option, but rather of identifying a set of "good" options, that may differ greatly in what criteria they optimize.

More relevant to our work is Multi-Criteria Decision Making (MCDM), a discipline that studies procedures to aid decision making when dealing with a large number of alternatives. Often, MCDM problems involve the ordering or classifying of alternatives [23], by imposing a procedure or strategy, like explicitly ordering or weighting criteria [24]. On the other hand, we use the term trade-off analysis to describe a process that is less structured: *the exploration, aided by visualizations, of multiple alternatives in order to understand the different criteria they optimize, without any imposed procedure or strategy*. Nevertheless, we are motivated by work in MCDM to identify possible tasks to support in a visualization tool for trade-off analysis.

Trade-off Analysis Tasks

Trade-off tasks feature in many visualization and visual analytics systems, in particular where users, via an interactive interface, have to make decisions according to multiple criteria. Some systems treat trade-off analysis as a single coarse task. For example, Sedlmair et al. [25] analyse tasks that users engage in when conducting visual parameter space analysis to find the best parameter combination given some objectives. Booshehrian et al. [26] make use of visualization to understand trade-offs of a set of pre-computed simulations. They describe a task within the analysis workflow of scientists working in fisheries management, where the goal is to quantify trade-offs between selected options, avoiding sensitive regions of the parameter space and those with high uncertainty. Other works break down trade-off analysis into smaller sub-tasks, often focusing on particular exploration strategies such as ranking of alternatives, or weights to express preferences. For instance, in the LineUp system [27] users can rank solutions and evaluate their performance relative to each other. The tool also provides means to combine dimensions (or attributes) and weights to convey priorities for the different attributes. Similarly, Weightlifter [28] helps analysts understand the effects of different weights for multi-criteria decision spaces with up to ten criteria.

Work that breaks trade-off analysis tasks into finer components or sub-tasks can also be found outside the visualization and visual analytics community. The work of Hakanen et al. [7] from operations research is the closest to ours. They built an interactive visual analytics systems where a decision maker explores Pareto front solutions that are computed, iteratively, based on their user preferences. They discuss trade-off analysis in the context of interactive multi-objective optimization, and identify seven high-level tasks: compare Pareto optimal solutions, specify preferences, check feasibility of preferences, determine most preferred solution, learn about problem characteristics, detect correlations, and post-process the most preferred solution. We hypothesize that most of their tasks [7] are still relevant for both general trade-off datastets (not just Pareto front solutions) and for a-posteriori methods such as ours, with the exception of the tasks "specify preferences" and "check feasibility of preferences" which are specific to interactive multi-objective optimization.

Visualization in Trade-off Analysis

Trade-off datasets are almost always multi-dimensional, covering objective dimensions to be optimized, and whose values vary in relation to other dimensions (or parameters). Many existing visualization techniques for multidimensional data have also been used for trade-off analysis, including tabular visualizations, parallel coordinates, scatterplot matrices, and self-organizing maps (for an overview of visualization techniques for trade-off sets see [29], [30]). Because trade-off exploration is a complex process [2], existing work looked at how to better support decision makers during the analysis, such as in terms of powerful interactive visualizations and robust ranking and optimization algorithms [28], [29], [31], [32], [33]. Tušar at al. [29], however, argue that trade-off sets (such as those found by evolutionary multi-objective optimization algorithms), require further support that traditional multidimensional visualizations do not necessarily cater for. Our study investigates these additional needs, and hypothesizes that provenance visualization can aid trade-off analysis.

Although trade-off analysis and provenance visualization are prominent areas of research, little work exists that looked at exploiting provenance information to specifically assist trade-off analysis. An exception is recent work by Cibulski et al. [34], who developed an interactive Pareto front visualization tool to explore multi-criteria alternatives in engineering design. Since their focus is on trade-off analysis, they describe 13 design requirements for this domain, such as showing criteria ranges for simulation steering, supporting selections, and highlighting conflicting criteria; but they also describe one additional requirement for provenance pertaining to storing favorite alternatives for future comparison. Trinkaus et al. [35] presented a generic multi-criteria decision support system called knowCube. They use provenance to support trade-off analysis by logging navigation paths, then clustering, storing and viewing "good" decision alternatives. As such, past navigation paths and good candidate solutions are stored in a log file, and the user can later replay them in an animated movie-like fashion. KnowCube provides a simple visualization in the form of a radar chart to show the different decision alternatives, and is demonstrated through real-life applications, but no user study or formal evaluation were conducted.

In summary, we were inspired by the recommendations of these trade-off analysis systems and provenance studies to design our own provenance-driven trade-off analysis probe and user study. To the best of our knowledge, no prior user study looked at the effect of provenance visualization on trade-off exploration.

3 VisProm: Provenance Visualization tool for Trade-off Analysis

We designed VisProm (Visualization Provenance master) to act as a technology probe [8]. Technology probes are systems traditionally used to field test the usage of a technology in real world settings, and inspire ideas for new technologies to support user needs. As such, VisProm had to be robust, but also effective for field use. Our goal was to see how provenance visualizations can aid trade-off analysis, focusing in particular on *in-visualization* views that are embedded in the exploration environment (rather than dedicated views). To this end we needed as a starting point a tool that has been shown to be appropriate for trade-off analysis, that we could then augment with in-visualization provenance views.

The inspiration for the basic trade-off support (before adding provenance) comes from EvoGraphDice [36], [37], [38]. This tool relies on a scatterplot matrix (SPLOM), that is well suited to communicate relationships between dimensions, such as correlations, and distributions of data points [39], [40], and is thus appropriate for visualising relationships between objectives to show trade-offs. The SPLOM is combined with multiple query selections, differentiated by colour, that help experts narrow their search space to important parameters [29]. Most importantly, EvoGraphDice has been used effectively in the past for trade-off analysis conducted by groups of agronomy experts [2].

While we did not use EvoGraphDice, in our new VisProm probe we replicated the following functionality from EvoGraphDice: (i) a scatterplot matrix of all dimensions from the dataset; (ii) a main scatterplot that is seen in larger detail; (iii) a query panel where users can choose colors in order to perform selections on the main scatterplot (visual query sculpting); (iv) a datatable available on demand, with points colored depending on the visual queries they belong to; and (v) a means to create new dimensions by combining existing ones using arithmetic operators and weights. This basic version also includes two components that can be considered as simple forms of provenance support: a history of actions such as view changes and queries, as well as a list of favorite views (or bookmarks) that is populated by the user.

We implemented the above functionality in VisProm, and also provided additional support designed explicitly for provenance tracking. Thus, although the trade-off analysis part of VisProm is inspired by EvoGraphDice, they are distinct: EvoGraphDice is a desktop application that uses an interactive evolutionary algorithm to guide user exploration, and does not include in-visualization provenance aside from a basic history and favorite views.

In the past, EvoGraphDice was used to study the role of expertise in collaborative exploration of complex model simulations (and how domain experts structure their exploration) [2]. Our current work goes deeper into the analysis tasks domain experts engage in when they explore trade-off datasets, whether they consult provenance views and how they use them while they are engaged in the exploration.

An informal test of EvoGraphdice [41] which included a provenance view as a separate window was not always consulted by users, who were often too absorbed by their main task. Thus in our approach we provided instead simple in-visualization views embedded in the analysis environment, in an attempt to make them always visible. We chose simple designs to not clutter and not compete with the main analysis views, but this begs the question of whether these simple designs are consulted and when, and if they carry enough information to support provenance needs. A contribution of our work is studying if the same, simple provenance designs can do both.

The basic trade-off support functionality and our provenance additions can be seen in Figure 1. The additional provenance designs were motivated by workshops with expert users.

3.1 Design process

To inform the design of VisProm we ran three brainstorming sessions with domain experts and visualization researchers (early results related to storytelling are published in a late-breaking-work report [41]). Nine participants took part in these sessions that lasted around two hours each. Five participants had design, HCI or visualization background and four were domain experts in agronomy. Participants were first shown a scatterplot matrix tool with basic provenance support (History & Favorites) [2], followed by an ideation and brainstorming phase on aspects that could help them keep track of their goals and revisit their exploration.

We report next the main recommendations derived both from our workshop participants,

as well as findings from previous work (section 2.2). These guided the design of VisProm.

[R1] History and Favorites (existing features) are important for reverting to past steps when errors occur and for storing important views. Thus they should be maintained.

[R2] Key objective dimensions that have been visited (or ones that have not) are important to know. This is confirmed by past work that observed that when experts have a-priori hypothesis they use objectives to prioritize their exploration [2].

[R3] Views already visited, i.e., what relationships between dimensions have been considered, was a common suggestion as a type of information the experts would like to track.

[R4] Combined dimensions they created and with what weights, were identified by some experts as important to track. Previous work has also discussed the importance of weights for expressing preferences [28].

[R5] What objective dimensions were maximized or minimized in their decisions, or more generally what range of the dimensions their decisions fell in, were also identified as important to keep track of. Past work also highlights the need to identify and “lock” part of decision ranges [35].

[R6] Alternative search spaces considered should be tracked. Our experts mentioned this in the workshop, and past work on trade-off analysis has identified that comparisons of alternatives are important [7].

[R7] A way to summarize the entire exploration would be desirable. This is an important topic in provenance visualization, not only for trade-offs (section 2.1).

3.2 Additional *in-visualization* Provenance Designs

We discuss next the different in-visualization components of VisProm that we added to communicate provenance, and how these support the requirements identified before.

This is the author version of the article. The final version is published in IEEE Transactions on Visualization and Computer Graphics 2022, at: 10.1109/TVCG.2022.31710745

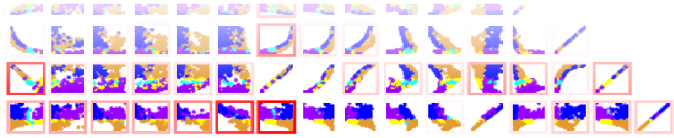


Fig. 2. ViewCount: in this partial view of the SPLOM we see the View-Counts as red borders around the thumbnails, indicating how frequently they have been visited (more saturated red for more visits).

3.2.1 ViewCount on Scatterplot matrix (Figure 2)

Our tool relies on a SPLOM matrix that displays thumbnails of scatterplots for all pairs of dimensions, coupled with a main view of one selected scatterplot shown in large scale that users can interact with. The SPLOM acts as a navigation tool, letting the user choose which pairs of dimensions they want to interact with in the main scatterplot view. Users can select points on the main scatterplot using colored queries, and see the impact of these selections in all thumbnail views. This ensures the user knows exactly where the selected points fall under different dimensions that they may be trying to optimize.

Each scatterplot in the SPLOM corresponds to a relationship between two dimensions, for example correlations or competing objectives that represent trade-offs - when one increases the other decreases. To help analysts keep track of what pairs of trade-offs they have already considered **[R3]**, we log how often they visited each scatterplot. We augment the SPLOM thumbnails with a colored edge, whose opacity changes in real time in proportion to the number of times the corresponding scatterplot was visited, relative to the most accessed scatterplot

3.2.2 DimensionCount table (Figure 3a)

Our participants in the workshop expressed the need to track the important objective dimensions they have already tackled in their analysis **[R2]**. To aid this task, we introduced a table with one cell per dimension, where each cell contains a bar that represents the number of times each dimension has been accessed (all bars are relative to the total access counts in the SPLOM). The length of the bars indicates which dimensions have been already considered through visits of scatterplots that include them, and which dimensions may have been overlooked. This representation differs from the ViewCount on the scatterplot matrix, as it focuses on the access history of individual dimensions rather than pairs. This design resembles past work, like the Lumos attributes panel [19] that shows frequency of interactions with different dimensions, to provide in their case awareness of biases in the exploration.

We initially planned to only add the DimensionCount table on top of our collapsible datatable (Figure 4). In the datatable each column is a dimension and each row a datapoint with corresponding numerical values, as well as the colored selections they belong to (colored rectangles). Given the existing organization by dimension columns, we added here the DimensionCount for each column (dimension). Feedback from experts indicated that while the datatable is extremely useful to confirm surprising findings and double check values (also observed in previous work [2]), it nevertheless remains collapsed and not visible during the majority of the exploration. We thus decided to replicate the DimensionCount table on top of the visual query selection panel as well, so as to ensure it is visible at all times (Figure 3a).

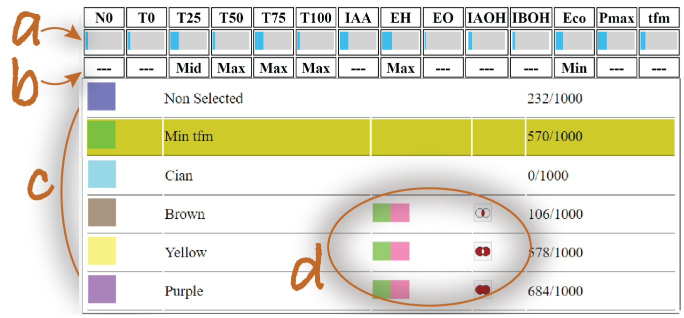
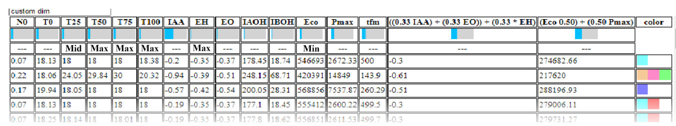


Fig. 3. The query selection panel includes: the dimension names and their DimensionCounts (a) as a bar that is filled depending on frequency; the SelectionRange tool (b) that indicates with text in what range (if any) the current selection falls under for the particular dimension; and finally (c) all queries/selections. Selections can be renamed (like the active selection *Min tfm*) and their colors changed. For selections that were created by group operations between other selections, their SelectionOrigin (original selections and operation) can be seen next to their name (d).



custom dim

N0	T0	T25	T50	T75	T100	IAA	EH	EO	IAOH	IBOH	Eco	Pmax	tfm	((0.33 IAA) + (0.33 EO)) + (0.33 * EH)	(Eco 0.50) + (0.50 Pmax)	color
---	---	Mid	Max	Max	Max	---	Max	---	---	---	Min	---	---	---	---	
0.07	18.13	18	18	18	18.38	-0.2	-0.35	-0.37	178.45	18.74	546693	2672.33	500	-0.3	274682.66	
0.22	18.06	24.05	29.84	30	20.32	-0.94	-0.39	-0.51	248.15	68.71	420391	14849	143.9	-0.61	217620	
0.17	19.94	18.05	18	18	18	-0.57	-0.42	-0.54	200.05	28.31	568856	7537.87	260.29	-0.51	288196.93	
0.07	18.13	18	18	18	18	-0.19	-0.35	-0.37	177.1	18.45	555412	2600.22	499.5	-0.3	279006.11	
0.07	18.25	18.14	18	18.01	18	-0.19	-0.35	-0.37	177.8	18.62	556851	2611.53	499.7	-0.3	279731.27	

Fig. 4. Top of the Datatable, including colored indications of points belonging to queries (right column). On the top just under the dimension names, are the DimensionCount and the SelectionRange tools.

3.2.3 SelectionRange table (Figure 3b)

Our experts commented that it is often challenging to keep track of what key objective dimensions are optimized (maximized or minimized) in the choices they make **[R5]**. In VisProm, choices are generally expressed as colored queries/selections of interesting points. To help analysts understand where their selections fall under the different dimensions, we added the SectionRange table. Each cell of the table corresponds to one data dimension, and indicates with text (Min/Max/Mid or —) two key pieces of information: (i) if the points in the current selection are mostly clustered under that dimension; and (ii) if this cluster tends to be in the max range of values of the dimension, the min, the middle range, or if no pattern is detected. We consider as clustered, datapoints that have a low value for the ratio `variance`/(`value range of dimension`). Then we determine the approximate position of the cluster in the range of values for the dimension (higher, middle, lower third). The SelectionRange table is placed below the DimensionCount table, and thus added to both the datatable and selection query panel.

Other work [19] visualizes where individual interactions with data-points take place along a dimension, compared to an ideal distribution for that particular dimension. In our case, our experts are more interested in larger zones of data-points (expressed as selections) and where these fall across all the different dimensions they are trying to optimize.

3.2.4 Selection Origin (Figure 3d)

Analysts can interact with the main scatterplot view and perform visual query selections (Figure 3c). The visual query panel helps them choose a color for each of their selections, and provides additional information regarding the number of selected data points. As in previous work [2], [37], we provide tools to sculpt the visual selections (adding and subtracting datapoints in different views). Past work [2] and our workshops indicated that experts express their search and trade-off criteria with these colored selections. They often compare them when they represent different decision alternatives **[R6]**, observing where they fall within the multi-dimensional space. One aspect that the experts requested was to also be able to perform grouping operations between multiple important selections (intersection, union, difference) in order to combine their criteria more effectively. In an effort to help them keep track of the provenance of such combinations, we added information about their origin in the form of two icons next to the query name: one with the grouping operation that created it and one with the colors of the original queries used in the grouping.

3.2.5 History and Favorites (Figure 1)

While not unique to VisProm, we mention History and Favorites as they represent basic provenance visualizations that are important for our participants **[R1]**. The basic history displays thumbnails of the last states of the tool - both selections and navigation changes. The Favorites acts as a bookmarking tool, where participants can explicitly store a view for easy access. Both History and Favorites are interactive, clicking on a thumbnail restores the saved view.

Our goal is to create in-visualization provenance views rather than summaries that can be considered outside the main exploration view. We thus do not provide elaborate summarizations of past exploration history **[R7]**. Nevertheless, our History provides a snapshot of the final workspace, that is valuable for remembering past analysis [42].

3.3 Implementation

As our work took place during the COVID-19 pandemic, we opted for a Web tool that could be operated easily without any installation and intervention from our part. We implemented VisProm using d3 and Node.js for the rendering aspects, and the Google Firebase database service to log the user actions during the exploration. For the individual thumbnails on the SPLOM matrix, we used Canvas drawing: as they are quite small, we optimized their rendering using sampling [43] that reduced the number of points drawn inside the thumbnails, but ensured the larger patterns are still visible. The prototype has been used on the Chrome browser with datasets of up to 20 dimensions and 1000 datapoints without any performance issues. The provenance visualizations on the SPLOM updates every time a selection is made or a combined dimension is created but will only update for every three accesses to the SPLOM thumbnails (this makes the exploration of the SPLOM smoother).

4 User Study Design

We conducted an observational study to understand how in-visualization provenance is used during trade-off analysis. We focused on the following research questions on *task support* and *usefulness* of provenance during and after the exploration:

[RQ1] What tasks experts engage in during trade-off analysis? How do provenance purposes manifest in trade-off analysis?
[RQ2] Does in-visualization provenance support *a-posteriori* analysis, such as by helping recall of past exploration?
[RQ3] Is in-visualization provenance taken into account by experts *during* trade-off analysis and how?
[RQ4] What are untapped opportunities for using provenance to support trade-off analysis?

TABLE 1
Datasets and participants (P1-5) for the three case studies. Each was conducted in two sessions, with a time interval in between (Weeks).

CS	Domain	#Dimensions	#Points	P#	Weeks
ml	Benchmark meta-analysis	10	3000	P1,2	1
eco	Sustainable farming	9	100	P2,3	17
w	Wine fermentation	14	1000	P4,5	156

4.1 Participants and Use Cases

Five domain experts (P1–5) from research institutions in France and the UK participated in this study (5 male, mean age 36.6). Participants were researchers (4 domain experts) and one first year PhD student. The recruitment procedure drew on previous collaborations between the authors and participants, or between participants, but none took part in the design brainstorming sessions described in section 3.1. At the time of the study, all participants were involved in different research projects, we call *Case Studies* (CS), where trade-off analysis is an important aspect of their work. Different domain experts participated in each case study, with the exception of participant P2 who was involved in two case studies (CS-ml and CS-eco).

Case Study 1: ML Benchmark Meta-Analysis (CS-ml): The goal of this exploration is to characterize possible trade-offs and relationships between different features of a number of benchmark Machine Learning (ML) datasets. An example *trade-off* concerns the number of features of a benchmark dataset, its intrinsic dimensionality (i.e., dimensions that exclude noise) and the average correlation between features. As the number of features grows, the intrinsic dimensionality is likely to grow. However, the feature average correlation is likely to decrease.

Case Study 2: Sustainable Framing (CS-eco): The goal of this exploration was to assess the trade-offs between different aspects of sustainability in farming practices in Europe. These sustainability aspects or indicators cover environmental, economic and social factors that are often in conflict, such as between fodder consumption, animal production and farmers' perceived happiness. An example *trade-off* pertains to animal production, which is likely to increase with fodder consumption, but in many cases has negative environmental impact.

Case Study 3: Wine Fermentation (CS-w): The goal of this exploration is to compare different wine fermentation strategies and their characteristics. In terms of *trade-offs*, experts are searching for fermentation strategies that maximize wine aroma (three esters), and minimize undesired compounds (higher alcohols) and the energy required to control fermentation.

All datasets were prepared by the domain experts themselves (see Table 1), who also brought their own research questions. We note that all domain experts performed basic statistical analysis of their data prior to this study, using software tools such as MS Excel, R or Matlab, but they have not used interactive visualization.

4.2 Study Procedure and Setup

Similar to Ragan et al.'s method [42] where participants conducted analysis in two steps one week apart, we run our study in two sessions separated by a time interval of one week for CS-ml, 17 weeks for CS-eco, and 156 weeks for CS-w. Since our focus is on studying provenance through the history of user interactions, in the first session we only collected history data, and we only activated provenance and analyzed the second session of each case study. We also note that for CS-w, whose first exploration session took place three years earlier, we used a similar trade-off analysis tool (EvoGraphDice [38]), which had all the basic functions provided by VisProm but lacked the new provenance visualization functionalities described in section 3.

In the first session, participants reached *new insights* in the form of: interesting relationships between aspects of the data (e.g., between animal production system typology and sustainability indicators, CS-eco); new findings about clusters of data points and how they are situated with regards to a decision boundary (CS-ml); or alternative exploration paths to reach the same (aroma) objectives (CS-w). To reach these insights, participants consulted multiple scatterplots, created numerous selections, and in some cases composed new combined dimensions.

All second sessions of the three case studies were run online using a video conference platform. Participants used their own machines with either a 13" or 15" monitor. For each case study, one participant led the exploration by interacting with the online tool and sharing their screen with the second participant (and the study facilitators) when they were not co-located with them (CS-ml and CS-w). Both domain experts equally participated in the trade-off exploration by proposing directions for the exploration, advancing new hypotheses and discussing findings.

Each session lasted $\sim$1.5 hours (mean 108 mins). The second session (analyzed) was structured in the following Parts:

1- Introduction [$\approx$10 mins]: Welcome and tool setup.
2- Reproduce Insight Without Log [$\approx$15 mins]: Our experts first saw the stripped down version of VisProm, *without* log data from their previous session. In this study part experts attempted to recreate the results they reached in their previous exploration relying on their memory. Our goal was to identify potential challenges in reproducing insights when no provenance history data is visualized.
3- Reproduce Insight With Log [$\approx$15 mins]: Our experts then saw the full version of VisProm with log data from their previous session visualized. We explained the additional provenance visualisations then participants were asked to reflect on their past exploration, in particular any aspects that they were unable to reproduce by memory. The goal of this study part was to observe how experts used the new provenance tools to revisit past findings.
4- Open Explore [$\approx$30 mins]: Participants continued their exploration with the full VisProm functionality, without any further instructions. The goal of this study part was to see if and how experts used the in-visualization provenance during their active exploration.
5- Discussion [$\approx$20 mins]: Finally, we conducted an open discussion session to elicit the experts' feedback about their experience, in particular related to how they used the provenance aspects of VisProm and what more they would like added or improved.

We chose to study provenance both when revisiting past analysis (days and even months later) and during active exploration. We suspect that the provenance needs differ: revisiting past analysis may require more support to acquire high-level understanding and reproduction of steps when memory may be faulty, whereas it may not be needed for reproduction of steps during active exploration and may even not be used in on-going analysis.

4.3 Data Collection and Video Coding

We collected video recordings of the three case studies (324 minutes), user interaction log files and observational notes. Two different authors of this paper independently coded each video using discrete events, then met to discuss annotation codes and resolved conflicts. We use the term event to refer to a chunk of video for which we attach a code pertaining to our research questions, mainly about trade-off tasks, provenance, and opportunities for provenance supported trade-off analysis.

We followed standard thematic analysis [44] in coding the videos and analyzing the collected data, using both inductive and deductive coding. We coded separately trade-off tasks and provenance events. For trade-off tasks, we did not rely solely on specific terms to find trade-off tasks (such as trade-off, objectives, criteria), we also observed what participants did and what they said to enrich the descriptions of those tasks. For example, one participants quote: *"this is NitOut it would be nice to combine with NiInput"* (CS-ml) was coded as "Combine interest zones & dimensions". Another event was based on our observation that the participant was checking where a selection fell with regards to the bisector (and later on confirmed in the Discussion). This was coded as the trade-off task "Locate & characterize". For provenance events and associated opportunities, an initial annotation scheme was decided before coding, based on the general provenance purposes proposed in Ragan et al. [4]. We coded events in the study parts where experts engaged with their data (Parts 2-4, 84.26%) as well as their comments in the discussion (Part 5, 15.73%).

In addition to coding for trade-off tasks and observed cases of provenance use, we also coded opportunities for provenance. We identified these when observing domain experts express hesitance in remembering important aspects of their exploration (current or previous), or when we thought the challenges encountered by the exploration may be overcome using provenance visualization. Our corpus contained 178 events in total (CS-w: 39.32%, CS-ml: 34.83%, CS-eco: 25.84%), related to trade-off tasks (25.84%), provenance tool use (23.03%), and opportunities (51.12%). A thematic saturation analysis [45] found conceptual stability approximately halfway through the data collection.

The results of the analysis are described in the following section. Participants comments are in *italics* and we indicate their case study, as well as whether they were translated using the † sign.

5 Results

We first discuss the trade-off tasks that our study participants engaged in **[RQ1]**. We report next a summary of the provenance purposes we identified, that refine ones from previous work [4] by adding sub-categories and concrete specifications from real-world trade-off analysis. We then show how in-visualization provenance was used both to revisit past analysis **[RQ2]** and during the actual exploration **[RQ3]**. We conclude with the opportunities that our probe helped us collect that can inform the design of future trade-off analysis systems that include provenance **[RQ4]**.

Further details about our study, including the video codes, participants' comments and analysis notebook, can all be found in `https://github.com/tradeoff-analysis/provenance`.

5.1 Trade-off analysis tasks [RQ1]

We coded key events related to the different tasks participants performed during trade-off analysis. Our goal was not to code each one of those events. Rather it was to first, showcase the variety of tasks decision makers engage in during trade-off analysis. And second, to investigate if there are new tasks not reported in past literature. We qualify these tasks as high-level, because we focus on user needs for trade-off support, rather than usage of any implemented feature of our system. As such, some of the trade-off tasks described below are not currently fully supported by VisProm.

Overall, there were 46 task events (CS-ml: 47.82%, CS-w: 32.60%, CS-eco: 19.56%), most of which were observed during the Reproduce Insight Without Log part where participants tried to reproduce an old insight without log data loaded into the tool (41.30%), and the *Open Explore* part where they conducted open exploration using the additional provenance support tools (52.17%). We only observed one trade-off task event in Reproduce Insight With Log as participants were mostly concerned with interpreting the loaded history data, and reflecting on how the previous insight was reached, rather than actively conducting trade-off analysis.

The authors of this paper coded these events and categorized them into the following *eight* high-level trade-off tasks:

[T1] Minimize & maximize (6.52%): When aspects of the multi-criteria problem were clearly described, participants took a direct approach to explore trade-offs by minimizing or maximizing known objective dimensions or criteria. For example, participants of CS-w made various colored selections to maximize the amount of aroma, and to minimize undesired wine compounds and the energy and time required to control the fermentation process.

[T2] Locate & characterize (28.26%): When little or no information is known about the structure or dependencies within the multi-criteria problem, participants tried first to locate interesting areas of the trade-off space using known thresholds, decision boundaries or ranges that are important for the problem domain. For example, experts in CS-ml examined the trade-off between the number of features and number of samples in their dataset, focusing on a cluster of points having 2-16 features. They then tried to characterize this cluster by checking its spread and distribution in other views. As the number of selections grew, participants gave meaningful names to their selections such as “grains high subs” (CS-eco) or “optimum” at the end of the exploration (CS-w).

[T3] Cascade & refine (15.21%): Participants deepened their understanding of the multi-criteria problem and its characteristics by *cascading* selections across multiple views, using brushing and linking. Oftentimes, they *refined* and narrowed down the initial selections, in order to find the best possible solutions. For example, in CS-w, starting from an initial large selection of fermentation recipes requiring a minimum level of initial nitrogen (149 recipes), participants found a much reduced set of “optimal” solutions (9 recipes), based on successive cascade and refine operations, using three criteria: aromas, fermentation time and energy.

[T4] Rank & prioritize (4.34%): A key decision participants often made when exploring trade-offs was to decide the order in which to examine the different objective dimensions. This order or ranking may have a different meaning depending on the use case or participants’ research questions. For CS-w, the order determined the size of selections (i.e., interest zones). Initial criteria were less strict and participants tended to make larger and more generous selections (e.g., on initial nitrogen amount), whereas selections based on subsequent criteria were more strict and thus their sizes were smaller (e.g., aromas). For CS-ml, the rank did not determine the size of selections. Instead, participants first visited features of their dataset that are most relevant to their research questions.

[T5] Create branches & compare alternatives (13.04%): When resuming previous analysis, participants wanted to branch out with a different but related goal, or create an alternative exploration path often with shared objective dimensions. New branches and alternative exploration paths helped domain experts discuss possible explanations and reasons behind certain trade-off choices.

[T6] Combine interest zones & dimensions (10.86%): Our participants often used a divide and conquer strategy, looking at selected zones of the trade-off space across different dimensions, and then combining them to explore more complex trade-offs. For example, participants created multiple selections to express different interest zones such as farms with high subsidies, and those who support organic practices (CS-eco). They then combined these selections to see farms that fell under both zones. Our experts were also interested in combining objectives, and dimensions more generally. In particular, they selected a subset of objective dimensions and assigned weights to reflect their relative importance.

[T7] Observe coverage (4.34%): Participants wanted to keep track of values explored within a specific data dimension or across many. They commented on whether they visited most dimensions, but sometimes they had difficulty remembering the dimensions they visited and the part of the exploration where this happened.

[T8] Find clusters, correlations & outliers (17.39%): Like in multidimensional data exploration, participants looked for clusters, correlations and outliers in their datasets, including between pairs or multiple conflicting objective dimensions. They explored groups of data points forming a cluster in one of the views, or meeting a specific criteria, then looked at how these relate to other dimensions.

In section 6.1 we discuss the differences between our identified tasks and these reported in past work on trade-off analysis in operations research [7], and reflect on the generalizability of these tasks. In the next section we will see how the additional provenance tools aided specific trade-off tasks, like keeping track of coverage, and what provenance needs they met.

5.2 Refined provenance purposes and objects

We next looked at how provenance purposes manifest within trade-off analysis. We report on aspects of provenance that we observed or were reported to us by our domain experts (74.15% of coded events). When analyzing these events, not only did we observe all of the provenance purposes discussed in [4] (Figure 5), but we also identified aspects central to the trade-off analysis process or the data records that participants were interested in. We call these aspects *objects of provenance purposes*, and in our coding, they were identified by asking ourselves this question: “**What** do our experts want to X here to support trade-off analysis?” Where X is a provenance purpose [4], either recall, replicate, recover, meta-analyze, or present. We categorize these provenance objects into the following eight types:

[O1] Interest zones (27.27%): Constraints that decision makers are willing to accept for one or multiple dimensions, such as a threshold, range, boundary, or distribution.

[O2] Exploration structure & steps (25%): The analytics steps taken during trade-off analysis, or the general phases and scenarios that constitute the exploration, which can include refinement, branching and alternative paths.

[O3] Focus & priority (15.90%): The dimensions that the experts focus on, either as single, multiple or combined dimensions.

[O4] Coverage (11.36%): The extent to which the trade-off space has been explored, such as in terms of data dimensions and views visited, or the data files loaded into the tool.

[O5] Data semantics & origin (8.33%): The meaning conveyed by the data dimensions, specific values or interest zones, and the context and data collection methods used to interpret them.

This is the author version of the article. The final version is published in IEEE Transactions on Visualization and Computer Graphics 2022, at: 10.1109/TVCG.2022.31710749

	Recall	Replicate	Recover	Meta	Present	#Events
Interest zones %	78	3	11	3	6	36
Exploration structure & steps %	48	27	15	9	-	33
Focus & priority %	95	5	-	-	-	21
Coverage %	100	-	-	-	-	15
Data semantics & origin %	100	-	-	-	-	11
Exploration approach %	-	-	-	100	-	10
Verification & validation %	75	25	-	-	-	4
Annotation %	-	-	-	-	100	2

Fig. 5. Our identified objects and how they are distributed across Provenance purposes [4] ("Present" purpose also includes Collaboration).

[O6] Exploration approach (7.57%): High-level reflections on the exploration strategy, which often includes a motivation, a rationale or an evaluation of the approach.

[O7] Verification & validation (3.03%): Confirmation of known insights or findings, such as to verify or validate a dataset.

[O8] Annotation (1.51%): Notes and labels to document important aspects of the exploration such as key views, useful insights, or the rationale behind the exploration.

Interest zones **[O1]** appears to be the most important provenance object in trade-off analysis as experts were interested in their recall, replication, recovery, meta-analysis and presentation (Figure 5). This is followed by exploration structure & steps **[O2]** (e.g., alternatives or branches), where experts were also interested in using this information for all purposes except for presentation, presumably because the trade-offs and insights expressed as interest zones, were more valuable to share with others than the exact steps to get there. Other objects were tied to specific purposes, such as recall for both coverage **[O4]** and data semantics & origin **[O5]**, meta-analysis for the exploration approach **[O6]**, and presentation for annotation objects **[O8]**.

Our experts were interested in remembering all provenance objects, apart from the exploration approach. This could be because experts did not experiment with different approaches to explore trade-offs given their brief exposure with our tool, and so they did not feel the need to be reminded. However, they still reflected on their choice of exploration approach, and sometimes also on their exploration structure & steps, and interest zones. They were mostly interested in *reproducing* past exploration structure & steps, and, to a lesser extent, past interest zones, the dimensions they focused on **[O3]**, and the results of any verification & validation they carried out **[O7]**. Finally, they mostly wanted to *recover* interest zones and their past exploration steps. Coincidentally, these are the two aspects they struggled with the most, either by accidentally loosing selections, or because the exploration had too many steps.

It is interesting to note that some objects have a direct mapping to the identified tasks. We discuss the importance of these identified objects for trade-off analysis design in section 6.2.

5.3 Provenance supports trade-off analysis [RQ2, RQ3]

We report next observed events of experts' using provenance views in trade-off analysis. These events are either direct observations of the experts using the provided tools based on pointing gestures or direct references, or cases where experts mention explicitly having used the tools. We rely on deictic and verbal references, as most of our provenance tools are non-interactive and thus we cannot log their use. The 41 events collected were distributed across case studies (CS-eco: 8, CS-ml: 13, CS-w: 20).

5.3.1 How provenance tools were utilized

We first report on how experts made use of theVisProm probe. Our goal in this section is to identify when the provenance views met the experts' provenance needs, sometimes in unexpected ways.

We explain how each provenance tool supports the tasks identified in section 5.1 (in **bold**) and the provenance purposes (in *italics*). A contribution of our work is introducing objects of provenance purposes, i.e., what the experts want to recall/replicate/annotate ... (section 5.2), so we report these explicitly (in *underlined*). For example in the sentence "X helped experts *Recall* details of their dimension *Coverage* **[T7: ObserveCoverage]**", the word Recall refers to the original provenance purpose [4], the notation [T7] is a trade-off task identified in section 5.1 related to dimension coverage, and the word Coverage refers to a refined object for provenance (Recall Coverage) related to trade-off analysis.

ViewCount that colors the SPLOM thumbnails based on use frequency, was the most used provenance tool (16/41). It aided *Recall* (13/16) and *Action Recovery* (3/16). In half the cases ViewCount helped experts keep track of *Coverage* **[T7]** in their exploration, i.e., objective dimensions they had already considered and ones they had not (7/16). They also used it as a reminder of the objective dimensions that were a *Priority* (**[T4:Rank&prioritize]** - 4/16). As one participant mentioned *"Because for example this [unvisited SPLOM thumbnail], I ignored completely."*, to which their colleague responds about priorities and unvisited regions *"Yeah but it's also a matter of priorities right? Maybe after you get all you can from these visualizations you might in the end go there [unvisited thumbnail] just to check."* (CS-eco). Although ViewCount does not provide temporal information, it sometimes (4/16) triggered experts' memory and helped retrace their *Exploration Steps*. When the red borders appeared, one expert that was struggling to remember their past analysis looked at the SPLOM and exclaimed *"Uh, now I remember something, we probably started from here"* (CS-ml).

History & Favorites followed in use (8/41). History was surprisingly used rarely (3/41), but always for *Action Recovery*. Participants used History thumbnails to recreate lost selections that represented *Interest zones* and trade-offs between objectives they had considered (1/3), and to retrace *Exploration Steps* (2/3). Favorites (5/41) were sometimes used for storing selections that corresponded to *Interest zones* (4/5) **[T2 & T3: Locate&characterize; Cascade&refine]** that participants wanted to share and discuss further, serving the broader purpose of *Presentation/Collaboration* (2/5); or for *Action Recovery* instead of the History in a few cases (2/5) to recover important lost selections. Experts also used them to reflect on the effectiveness of their past exploration (1/5) (*Meta-Analysis*).

SelectionRange presents in text the range selected points fall under each dimension (Min/Mid/Max/-), and came next in terms of use (4/41). Participants used it mainly to *Recall* (3/4) if the datapoints they were considering were in specific zones of the objective dimensions (*Interest zone* **[T1:Minimize&maximize]**). For example, when two experts were discussing a selection, one showed the SelectionRange bar to point out the solutions considered: *"Since we did not want to use a lot of energy, we have Eco in the mid-range [points to the SelectionRange]"* (CS-w†). In one event (1/4), our experts used the range bar to *Replicate* a past action, such as when they remembered their past strategy of giving *Priority* **[T4:Rank&prioritize]** to some criteria: *"So we try to find sub-groups of datasets that are like, small number of features [referring to SelectionRange]"* (CS-ml).

DimensionCount bars indicate how often a dimension was viewed. It was used less (4/41), in all cases to *Recall* information. Participants used it to keep track of *Coverage* **[T7:ObserveCoverage]** of the important objective dimensions they wanted to optimize (3/4), and to *Validate* that they remembered correctly their past exploration (1/4).

Queries allowed experts to make selections that express choices about datapoints that fit specific *Interest zones* **[T2:Locate&characterize]**. It was used extensively during exploration as experts explored alternatives **[T5:CreateBranches&Compare Alternatives]**, but we observed a few events where it was explicitly used to support provenance (4/41), in particular to retrace *Data Semantics* (3/4). The visuals in the **SelectionOrigin** of whether they were created from group operations on other selections, helped in one event experts *Recall* the meaning of a selection (1/4). In another, changing the name and color of their query (e.g., call a selection *"optimum"*, CS-w) helped them at a later stage to *Recall* what selections represented a good subset and what they were optimizing (2/4). In one event (1/4) seeing the number of datapoints in a selection allowed them to reflect on their exploration (*Meta Analysis*) and comment *"yes, here instead of selecting 73 points for the green selection [...] as we did in the past exploration we just selected 3 or 4 points, it's the same idea in today's exploration we drew a line on which we aligned points [...]"* (CS-w†).

Our participants also mentioned verbally (5/41) that loading and viewing their past history with the different provenance tools helped their analysis. They did not single out specific tools, but commented in the Discussion (Part 5 of the study) on how the visualizations holistically helped them retrace their past *Exploration Steps* (1/5) for *Action Recovery*; allowed them to *Recall* what they *Focus*ed on in past explorations in terms of dimension coverage **[T7:ObserveCoverage]** and also the combined-dimensions they created **[T6:CombineInterestZones&dimensions]** (2/5); helped them *Recall* selections that represented *Interest zones* **[T1,T2: Minimize&maximize; Locate&characterize]** (1/5); and enabled them to reflect (*Meta-Analysis*) on the effectiveness of their *Exploration Approach* (1/5).

Overall, we observed that our tools were used to help experts:
(i) recall important areas in the trade-off space in the form of selections (*Interest zones* **[T2:Locate&characterize]**), and what these selections represented and optimized (*Semantics* **[T1:Minimize&maximize]**). In total 13/41 events (SelectionRange, History, Favorites, Query);
(ii) keep track of the *Coverage* **[T7:ObserveCoverage]** of objective dimensions they had already visited and what remained. In total 10/41 events (ViewCount, DimensionCount); and
(iii) recall what objectives were of high *Priority* **[T4:Rank&prioritize]** (7/41 events - ViewCount, SelectionRange);
(iv) retrace past *Exploration Steps* (7/41 - ViewCount, History);
(v) or to *Validate* and reflect on the effectiveness of their *Exploration Approach* (4/41 - DimensionCount, Query).

5.3.2 Provenance at different exploration parts [RQ2,RQ3]

To understand if the provenance tools were used to recall past exploration **[RQ2]** or to aid the current analysis **[RQ3]**, we focus respectively on events that happened during the Reproduce Insight With Log Part and the Open Explore Part (28/41). The remaining events come either from the Discussion or before the logs were loaded in the provenance tools (when only Favorites and History were available) and are not discussed further in this section.

Interestingly, the provenance tools were used actively both during the Reproduce Insight With Log when they were revisiting past explorations (17/41 events), and during the active Open Explore (11/41 events). The longer Open Explore (30 min) had fewer events than the Reproduce Insight With Log (15 min), but this can be expected, as in the latter the participant's goal was to reflect and reproduce their past analysis - so provenance was at the forefront; whereas in Open Explore their primary goal was to explore new aspects of the dataset. Thus we are more interested in trends of use between the two parts, rather than a direct comparison of numbers. To that end we provide percentages of events within each exploration Part.

[RQ2] In Reproduce Insight With Log where participants were trying to reconstruct their past analysis, they mainly used the ViewCount tool (41.2%), followed by DimensionCount (23.5%) and Favorites (11.8%). We also observed use of the Query tool to reflect on the size of a past selection (5.9%). The remaining events are general comments about how the tools helped them to revisit past findings, but without specifying which tools. We did not observe use of the SelectionRange, the SelectionOrigin, nor of the History tool.

Unsurprising, most events relate to the general provenance purpose [4] of *Recall*-ing the past exploration (70.6%), with a few events where experts reflected on the quality of past exploration (*Meta Analysis*) (17.6%), or attempted *Action Recovery* (11.8%). When looking at the more specific provenance objects, participants mainly used the tools to track the *Coverage* of dimensions they had visited in the past exploration (35.3%) and to remember which ones were their *Focus & Priorities* (23.5%). The other provenance objects were less represented in this study part, with only a few events where participants tried to *Validate* (5.9%) or reflect on their *Exploration approach* (17.6%), retrace detailed *Exploration Structure & Steps* (11.8%) or remember specific *Interest zones* (5.9%).

Overall, we confirmed that participants used the in-visualization provenance when revisiting past findings. Based on the provenance objects they focused on, high-level summaries of what was covered and what were the priorities in the past exploration, are key.

[RQ3] The use patterns differ in Open Explore, with tool use being more distributed. ViewCount and SelectionRange were the most used (27.3% each), followed by History (18.2%), SelectionOrigin and Favorites (9.1% each). The few remaining events are general comments about their usefulness without specifying which tools. We did not observe any use of the DimensionCount tool.

These events point out to a different distribution of Ragan et al. [4] provenance purposes. While *Recall* remains important (45.5%), the main purpose is now *Action Recovery* (54.5%). The reasons why participants used the tools also differed. More than half the time it was to keep track of *Interest zones* (54.5%), followed by retracing their *Exploration Structure & Steps* (27.3%). Less frequently, they used it to verify *Coverage* and *Data semantics* (9.1% each).

Overall, we confirmed that experts also used in-visualization provenance while conducting a traditional exploration analysis. Based on the provenance objects they focused on, here participants were mainly interested in tracking more detailed information, such as keeping a trace of specific areas in the trade-off space that were of interest to them and their characteristics, as well as their analysis steps. Nevertheless, we did observe events where higher level goals, such as keeping track of dimension coverage, were met.

5.4 Opportunities for provenance visualization [RQ4]

Our goal in using VisProm as a technology probe was to identify situations were new designs could aid users in their tasks. We coded several events across use cases that presented opportunities for provenance (CS-w: 35, CS-eco: 29, CS-ml: 27). We chose to report such events as they provide concrete and actionable opportunities for provenance in trade-off analysis.

We found opportunities in Study Part 2 where experts were tasked to replicate past events **without** the additional aids of log data and provenance visualization (31.86%), but also in Reproduce Insight With Log (27.47%), Open Explore (24.17%), and the Discussion (16.48%). These provenance opportunities are analytical in nature (93.4%), although we also describe opportunities for data provenance related to data semantics and origin (6.6%). Below, we report on these opportunities using the general provenance purposes in [4]: recall (**73.62%**), replication (**12.08%**), meta-analysis (**12.08%**), and presentation & collaboration (**2.19%**).

5.4.1 Recall Opportunities

We observed more opportunities for provenance *Recall* than any other type, as oftentimes experts struggled to remember previous interest zones (35.82%), focus & priority (20.89%), exploration structure & steps (20.89%), data semantics & origin (11.94%), coverage (7.46%), and verification & validation (2.98%).

Recall ***Interest zones***: When trying to recall previous interest zones, expressed in VisProm as colored selections, experts discussed but often struggled to fully remember what they saw as important trade-offs and their characteristics. We observed opportunities for tracking and visualizing provenance of: **(i)** the view/s in which the interest zone (or selection) was defined; **(ii)** selection size, range and distribution; **(iii)** the thresholds and cut-off points they used to constrain the selection; **(iv)** when relevant the decision boundary they used to interpret the selection, for instance, whether the data points are above or below the bisector; **(v)** the nature of the trade-off selection, e.g., minimization versus maximization. In addition, experts also raised the need for tracking contextual information **(vi)**: for example to differentiate between past selections that express an insight from the intermediate ones that led to it; sometimes even remembering the research questions or hypotheses that motivated them to create selections can be hard. Interestingly, experts used unexpected cues to help them recall the reasons behind selections such as green color for preferable solutions, and red for undesirable or costly ones.

Opportunities. *Creating and refining interest zones (e.g., via selections) is crucial in trade-off analysis, as it helps analysts progressively learn the multi-criteria problem characteristics, determine most preferred solutions and compare alternative ones. However, experts struggled to remember key aspects of those selections not only when revisiting them after a period of time, but also during the same exploration session. We need to enrich trade-off selections with summary information and visual cues to help analysts quickly understand or recall the nature of the trade-off, and the rationale that led to create those selections (e.g., to minimize or maximize certain objectives)* **[T2:Locate&characterize]**. *Color is an interesting visual encoding for provenance as it appears experts assign meaning to hue and are sometimes able to remember this meaning without access to labels. While in some cases experts were able to reconstruct what objectives were maximized, what selections were important and fell within specific boundaries or had specific patterns, we believe there is an opportunity to help them track this more explicitly. For example, we can allow them to set thresholds that the system can keep track of (beyond min and max), and let them express what patterns are of interest (data above bisector, or of a specific distribution) and why, and allow them to differentiate between selections that are intermediate steps in their analysis and ones that express important findings* **[T1,T2: Minimize&maximize, Locate&characterize]**.

Recall ***Focus & priority***: Domain experts often mentioned some objectives are more important than others and this drove their exploration strategy. They sometimes reported they could not remember all of the objectives that they tried to optimize, but VisProm (DimensionCount, ViewCount) helped them with recall. We observed that participants were likely to remember plots they found interesting, such as in terms of their dimensions and shape, but not so accurately or easily the steps they took to get there. Moreover, sometimes participants were also able to guess why they did not focus on a particular dimension or view.

Opportunities. *During trade-off exploration, analysts need to recall what objectives drove the exploration and their priority* **[T4:Rank&prioritize]**. *Showing the individual or combined criteria they created, and the interesting views they visited (even high-lighting those that are not interesting) can help them infer their overall trade-off exploration approach. Besides communicating focus and priority, analysts can themselves express these, possibly visually to indicate objectives that have not been explored yet* **[T7:ObserveCoverage]**. *It is interesting to note that we observed participants attempting to keep track of priorities both when revisiting past exploration, but also during the time they were actively exploring to help strategize and plan their next actions.*

Recall ***Exploration structure & steps***: Participants had difficulty remembering the successive steps they took during the analytical process. Experts from CS-w also struggled to recall the first step that triggered a particular exploration path, and this was important as the order in which they visited the trade-offs mattered to them. When reasoning in terms of stages of exploration [2] rather than individual steps, participants were able to remember more, such as the part of the exploration where they analyzed an important trade-off, or when they found an important insight.

Opportunities. *Overall participants were able to remember phases of the exploration more than they did for steps, but not all stages were remembered indicating that different parts of the exploration may require different support. Besides showing phases of exploration as part of provenance, trade-off exploration might benefit from explicitly showing alternative or branching paths taken during the analysis, and emphasizing similarities and differences, as analysts are interested in making comparisons* **[T5:CreateBranches&CompareAlternatives]**.

Recall ***Data semantics & origin***: We had several events where participants failed to remember the semantics of the data they were exploring. This happened at three different granularities: **(i) Dataset:** we had events where participants mentioned explicitly the origin of data and relations that were not captured in the dataset; **(ii) Dimension:** In some cases experts struggled to remember what each dimension of their dataset captured. This is particularly important for trade-off analysis, as often experts combine dimensions that express objectives that are related (e.g., in CS-eco one expert had to reflect before confirming that a dimension in the dataset combined two others, and in CS-w another expert mentioned how remembering this is important); **(iii) Value:** We also observed events where experts were attempting to optimize dimensions, but could not remember if the objective needed to be maximized or minimized. There were also cases where they had trouble remembering the meaning of specific values or ranges.

Opportunities. *We need to provide data provenance at different levels - not only to enable experts to keep track of the origin of data, but also to capture and be able to retrieve the semantics of dimensions and values. This is particularly important for trade-off analysis, where some dimensions may capture more than one objective, and where the experts need to track the direction (range) that optimizes specific objectives* **[T6:CombineInterestZones&dimensions]**.

Recall *Coverage*: Participants had difficulties keeping track of data dimension coverage when they had limited provenance support (Reproduce Insight Without Log). However, when we provided them with provenance visualization tools, they commented on how these aided them in their task. When recalling past exploration, they also wanted to know whether the dimensions they visited came from single or multiple sources (data files).

Opportunities. *Given the high dimensional nature of trade-off spaces, there are opportunities to improve the exploration by keeping track, as the analysis progresses, of what data dimensions or sources have been explored, and highlighting areas that are under-explored. This can potentially show an implicit exploration bias or an intended prioritisation in analysing trade-offs* **[T7:ObserveCoverage]**.

Recall *Verification & validation*: Our experts were also interested in finding out whether important known insights were confirmed in their previous (and current) analysis, thus validating the approach which was used to generate the trade-off dataset.

Opportunities. *Provenance visualization could make it clear whether the explored dataset has gone through a validation procedure (whether manual or automatic). This can give analysts confidence in the data and avoids them repeating the verification & validation procedure in any future explorations.*

5.4.2 Replication Opportunities

The provenance opportunities we observed to support *Replication* during trade-off analysis are mostly related to Exploration structure & steps (81.81%). Experts attempted to reproduce sequences of events in order to replicate an insight or a lost selection. The replication of previous analysis could be an important step when resuming work carried out a long time ago. We found that experts might want to repeat previous analysis from the beginning to refresh their memory, or to gain more confidence in their findings, but often with variations, such as by examining additional dimensions or a new dataset. This could occur in a branching or refining manner. Another sub-goal of replication is to compare alternatives. For example, experts were interested in comparing alternative solution sets in terms of differences and similarities.

Opportunities. *There are opportunities to support replication during trade-off analysis by assisting analysts in resuming major or key phases of their exploration, including important selections and dimension groupings. Different types of replications should be supported, whether to continue the same analysis, or to create a new branch or a refinement from an old one* **[T3,T5:Cascade&refine, CreateBranches&CompareAlternatives]**.

5.4.3 Meta-Analysis Opportunities

Participants reviewed the analysis process, often to understand their *Exploration approach* and structure (90.90%), or to reflect on their choice of *Interest zones*. The majority of those opportunities relate to events that occurred during Parts 3 and 4 of the study (72.72%), both with provenance visualization tools enabled. For example, participants in CS-ml wondered whether the next step would be a fresh start, or a new branching session extending from the previous analysis. In CS-eco, when we asked participants to reflect on why they wanted to keep the provenance history, even though they were starting the open exploration (Part 4), they stated that previous analysis could provide new inspiration or good starting points.

Opportunities. *Automatic provenance tools can support meta-analysis, for instance, by showing similarities and differences between exploration sessions. They may provide useful information and metrics, to help analysts reflect more deeply on their trade-off analysis choices. This in turn, may provide inspiration and valuable starting points for the following steps of the exploration.*

5.4.4 Presentation & Collaboration Opportunities

The views experts visit can become complex over time as they create combined dimensions and new selections (e.g., using intersection). They might want to *Annotate* such views or take notes, to facilitate revisiting and sharing. Whereas one domain expert noted that it could be useful to directly annotate interesting views, another expert found annotations a burden and add complexity to the tool. The question of annotation was also raised in CS-w where domain experts came to our study prepared with notes describing the research questions and hypotheses they wanted to test.

Opportunities. *The role of annotations as a provenance tool for trade-off analysis merits further investigation, as our observations show that annotation support could help analysts address the growing complexity of views as the analysis progresses, but with caution as to not distract from the main trade-off task.*

6 Discussion

The use of our provenance-augmented probe by three groups of domain experts, allowed us to study trade-off analysis, and to distill new findings with regards to the tasks people engage in during trade-off analysis, and how provenance can support these tasks both when conducting a-posteriori and active analysis. Moreover, it helped us identify opportunities on how to better support such analysis tasks.

6.1 Trade-off analysis tasks

In section 2.2 we theorized that five of the tasks presented in Hakanen et al. [7] are relevant to our work, even though unlike their approach, we do not integrate an interactive multi-objective optimization procedure. While this is generally true, we went beyond this past work. First, we found a case where our tasks are more refined and detailed, and second we identified additional trade-off analysis tasks not reported in [7]. We contributed a set of refined tasks that fall under their *determine most preferred solution: minimize & maximize*, *rank & prioritize*, and *combine interest zones & dimensions*, which are all tasks our participants followed to narrow down and converge to their preferred solutions. We also further contribute two new tasks: *cascade & refine* and *observe coverage*, that our participants found to be important for trade-off analysis.

Their other tasks have a direct correspondence to the ones we identified. Their task *compare Pareto optimal solutions* corresponds to our more general task *create branches & compare alternatives*. Their *learn about problem characteristics* and *detect correlations* tasks directly map to our *locate & characterize* and *find clusters, correlations & outliers* tasks respectively. Thus, together with the refinement of the *determine most preferred solution* task, we independently identified several tasks reported in operations research work [7]. This replication indicates our technology probe was effective in aiding us to identify high-level tasks that are generalizable across tools and research domains. The *post-processing* task is beyond the scope of our study, as we did not do a follow-up with our participants to investigate whether the insights gained from the exploration session helped them conduct new experiments or generate new trade-off datasets.

Our trade-off tasks of finding clusters, correlations & outliers, also correspond directly to high-level tasks in traditional analytic task taxonomies, such as the tasks identified by Amar et al. [46] on finding anomalies, clustering and correlating. Other high-level trade-off tasks like observing coverage and defining focus & priorities (that are key in trade-off analysis), do not have a direct correspondence. This is also the case for trade-off tasks such as creating branches & comparing alternatives, looking for interest zones that maximize/minimize objectives, and characterizing interest zones. These different tasks remain fairly high-level and do not have a direct equivalent in analytic tasks; nevertheless, in order to achieve these trade-off tasks, our experts engage in a combination of low-level analytic activities [46], such as deriving threshold values, filtering, sorting, characterizing distributions, finding extremums, etc.

While scatterplot matrix tools are popular in trade-off analysis visualization [29], [30], it is natural to consider if these refined and new tasks may be specific to the scatterplot matrix design of our probe, that favors sculptured selections. We believe the majority of our identified tasks are generalizable and not specific to our probe design. A scatterplot matrix view favors the identification of relationships between two dimensions at a time (e.g., correlation [40]). Nevertheless, our tasks relate either to tracking progress and prioritizing (rank and prioritize dimensions, track coverage of multiple dimensions), or to understanding possible solutions in the search space across multiple dimensions (e.g., identifying possible solutions that minimize or maximize a dimension of interest, and characterize zones of interest across multiple dimensions).

It is possible that other tasks, such as cascading & refining of interest zones, may indeed be influenced by how analysts select regions of interest in our tool, using sculptured queries; or the combine multiple dimensions task may be due to the fact that a single scatterplot can only show two dimensions at a time. For the latter, we feel combining dimensions makes sense irrespective of visualization, our experts often had dimensions (like aromas) that they semantically grouped together and wanted to treat as a group.

6.2 Provenance objects in trade-off analysis

We identified eight provenance objects that characterize provenance purposes from previous work [4]. These are objects that our experts wanted to recall, replicate, recover, reflect upon or present & communicate. Out of these objects, we found that interest zones and exploration structure & steps are the most important, as our domain experts frequently used or mentioned them during their exploration. They should thus be supported by provenance visualizations in future trade-off analysis tools.

Our provenance objects do not all reside on the same analytical or abstraction layer [6], [47]. Some are directly related to exploration actions [47] and thus can be tracked using low-level user interactions such as view and query selections (interest zones, exploration structure & steps, focus & priority, coverage). Other provenance objects are situated at a higher level pertaining to insight actions (data semantics & origin, annotation, verification & validation), or meta-actions (exploration approach). Tracking and visualizing provenance data related to exploration actions is more straightforward and is easily supported in interactive systems, whereas provenance for higher abstraction layers is more challenging to capture and visualize, although progress is being made for example in automated data insights [48]. This highlights the need for further research in detecting and visually summarizing insight actions and exploration structure to support trade-off analysis.

Provenance objects express ‘what’ aspects of the analysis experts need support with (to recall/replicate/etc), and were observed either through the use of the provenance tools and experts’ discussions (section 5.3.1), or observed opportunities (section 5.4). Some of these provenance objects directly map to, and support, identified trade-off tasks (section 6.1). For example the direct mapping between the Focus & priority object and Rank & prioritize task, and between the Coverage object and the Observe coverage trade-off task. We also observed analysts recall/replicate/reflect upon other objects, namely interest zones, especially for trade-off tasks such as Minimize & maximize and Create branches & compare alternatives. Or the ’Exploration structure & steps’ object that often appeared as an opportunity, expressing the need to keep track of past branches and alternatives for the task Create branches & compare alternatives. These tasks of Minimize & maximize and Create branches & compare alternatives relate to trade-off tasks from previous work [7]: determine most preferred solution and Compare Pareto optimal solutions respectively.

We thus feel the objects ‘Focus & Priority’, ‘Coverage’, ‘Interest zones’ and ‘Exploration structure & steps (to create branches and compare alternatives) are central to trade-off analysis and require further exploration in order to design provenance views that are well-adapted to such analysis. The remaining objects we observed (Exploration structure & steps for action recovery rather than branches and alternatives, Data semantics & origin, Exploration approach, Verification & Validation, and Annotation) could be relevant to any type of visual analysis exploration. Future work may also identify more trade-off tasks and related objects.

The nature of our study (revisiting past findings and active exploration), naturally relies on our participants remembering their past analysis, and thus focuses on replication or recovery of past results or choices. Thus our findings often link provenance objects with the provenance goals of Recall or Replication (section 5.2). We see these same objects (e.g., Interest zones) also appear with other purposes such as ‘Recover’ and ‘Present, so we expect these objects are applicable across provenance goals. Nevertheless, further studies are needed to investigate whether different types of analysis, focusing less on recall and replication, and more on reflection and meta-analysis, could clarify which objects are important for which provenance purposes, or even reveal other types of provenance objects associated with other purposes.

6.3 How in-visualization views support trade-off tasks and provenance objects: during *a-posteriori* analysis and *ongoing* exploration

Most work evaluating provenance considers provenance itself as the main user goal, for example to understand and recall past exploration. Fewer consider the impact of provenance views during the exploration itself (see section 2.1). We consider both together, in a structured evaluation of experts conducting real-world analysis.

This poses a challenging question of design balance: as data exploration is the focus of our experts, can we use provenance to support both a-posteriori analysis and active exploration? Motivated by past informal observations that experts largely ignored separate provenance views, we decided to integrate in the analysis environment simple designs that do not add visual complexity but are always accessible. We set out to see if and how these designs were used, both a-posteriori and during the analysis, and how they tied to trade-off needs.

Our findings show that our domain experts found in-visualization provenance views useful both when revisiting old analyses and when starting new ones.

In *a-posteriori exploration,* when revisiting old analysis, domain experts were able to first confirm past findings in the use of traditional views, such as History and Favorites. The use of a stripped down History that only showed a final snapshot of last exploration, and Favorites were a great aid for recall. Our experts explained that they preferred these few snapshots rather than detailed step-by-step recollection of previous analysis. Other studies investigated what granularity or level of detail of provenance information is appropriate for different tasks [42], [49], [50]. Most relevant to our work is a user study by Ragan et al. [42] who found that even lightweight provenance visualizations (such as ours), including simple in-visualization cues and low fidelity snapshots of final workspaces or important views that contain an insight, can be beneficial for recall and helping the process memory.

We were also able to extract patterns of use from our new in-visualization views during this a-posteriori exploration. Our experts used provenance views that mainly communicated high-level information, such as views with counts of dimensions explored or scatterplots visited. In particular, we identified that these views aided experts mainly to recall or reflect on: their coverage of the objective dimensions, their focus and priorities in terms of objective dimensions in their past analysis, and to a lesser extend their past exploration strategy. As mentioned, understanding coverage and focus & priority were identified both by us, but also previous work [7], as critical for trade-off analysis irrespective of the tool used. As such, this finding highlights the importance of providing provenance support for them when a-posteriori analysis is expected.

Whereas in the *ongoing* analysis, experts used our in-visualization views focusing on more detailed information. For example, they used them to keep a trace of specific areas in the trade-off space that were of interest to them and their characteristics, as well as their analysis steps. Our experts often consulted what range their selections fall in in the trade-off space, or what selections were grouped to generate new ones. These point to the need to support recall exploration steps, but most importantly the recall of aspects related to interest zones (one of the trade-off provenance objects identified in our work). Aspects that may be interesting to recall relate to several trade-off tasks that are associated with characterizing preferred solutions or interest zones: for example if the interest zones maximize or minimize objectives, if they fall under specific thresholds, if and how they have been refined, if they represent an alternative branch, etc.

As our observations of ongoing analysis stem from a scatterplot matrix probe, it is possible that the importance of interest zones may be influenced by the existence of visual sculptured query selection. However, as discussed earlier, the task of ‘determine preferred solution’ [7] that optimizes some dimensions is very general in trade-off analysis. And the characterization of these possible solutions or areas of interest (e.g., where they fall across other important dimensions) includes simultaneously maintaining awareness of multiple dimensions and thus is very likely applicable to any multi-dimensional visualization tool.

Provenance is a secondary task supplementing data exploration, but we observed many cases where our visually simple and in-visualization provenance tools were used effectively, both when reconstructing past trade-off analysis and during an open-ended analysis. More work is needed to investigate whether provenance support needs to adapt to more specific and refined stages of these analysis. For example, during an active exploration, we can expect that the start of the analysis will be similar to a-posteriori analysis: verifying and validating past knowledge, checking coverage of dimensions and setting priorities, in order to ground the upcoming exploration. Whereas in the middle of a trade-off analysis exploration, it is possible that the main considerations are keeping track of alternatives or of characterizing the identified interest zones. And towards the end of an exploration, when experts are close to converging and sharing of results, it is possible that tracking coverage and reflecting on the exploration strategy may once again become important. More studies are needed to compare different design choices for provenance as a secondary task, including understanding better potential exploration phases [41], what provenance objects are important in these more refined phases, and how to draw user’s attention and awareness [51] to the visual cues when the information is needed.

6.4 Trade-off specific opportunities

The use of provenance tools by our study participants helped us identify different types of opportunities for recall, replication, meta-analysis, representation and collaboration. These opportunities are directly related to actual challenges our study participants faced during trade-off analysis, for which we thought provenance could help overcome. The general provenance opportunities that are discussed in the literature confirm some of the provenance needs we identified for trade-off analysis, such as the recommendation for capturing provenance information with varying levels of detail [4] and the associated opportunities for multi-layer provenance [14]. The overlap between our recommendations and those in the provenance literature is expected, as trade-off analysis is a subset of and a specific type of data exploration.

Moreover, we covered new ground with regards to the provenance objects associated with the identified trade-off tasks. Our opportunities for helping analysis recall/replicate/reflect indicate these act upon objects that can inspire future designs, focusing on: interest zones, exploration structure and steps (like refining and comparing alternatives), focus & priority, and coverage. In particular, our call for supporting provenance capture, recall and replication of focus & priority argues for storing and exploiting information highly relevant to trade-off analysis such as the objectives users focus on, and how these are prioritized, which is intrinsic to trade-off analysis, but less pertinent in general data exploration and provenance.

7 Conclusion and Limitations

We conducted an observational study with three groups of domain experts who used a provenance-augmented probe to explore their own trade-off datasets. Our aim was to investigate whether and how in-visualization provenance can support real-world trade-off analysis. Our findings show that domain experts engage in different high-level tasks that are key in trade-off analysis, such as observing coverage, defining focus & priorities, branching alternatives, and looking for interest zones that maximize/minimize specific objectives. These tasks were supported by our in-visualization provenance views, which helped our experts recall, replicate, recover, reflect upon or present & communicate different aspects of their exploration, that we identified as trade-off provenance objects. From these, keeping track of coverage and priorities, and to a lesser extent exploration strategy, was the most important provenance aspect (or object) for our experts when revisiting past trade-off analysis; whereas interest zones and exploration structure & steps were crucial when starting new trade-off explorations.

While the use of the provenance probe helped us identify opportunities and challenges, it may have also influenced our findings. Our participants had access to a specific visual representation (SPLOM) and in-visualization provenance views, which may have focused their feedback to tool-specific issues. We believe the high-level findings, such as many of the identified trade-off tasks, provenance objects and opportunities hold irrespective of the underlying analysis tool, nevertheless it remains future work to replicate them in different trade-off analysis environments. Our probe was also used by experts from different domains and we report our findings collectively. This we hope provides a broad coverage of user needs when it comes to trade-off analysis, but it does not help us understand if the nature of provenance requirements is common across experts. Future work should consider if domain plays a role in both trade-off analysis tasks and provenance needs. Finally, we looked at pairs of analysts exploring trade-offs simultaneously. It is possible different provenance objects and even tasks emerge when analysts work asynchronously, for example during hand-off or when establishing common ground.

Acknowledgments

The authors would like to thank our domain experts for participating in this study, and for their feedback on VisProm.

References

[1] M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer, and T. Mller, "Visual parameter space analysis: A conceptual framework," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 20, no. 12, pp. 2161–2170, 2014.

[2] N. Boukhelifa, A. Bezerianos, I. Cristian Trelea, N. Perrot, and E. Lutton, "An Exploratory Study on Visual Exploration of Model Simulations by Multiple Types of Experts," in *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, CHI '19*, 2019. [Online]. Available: https://hal.inria.fr/hal-02005699

[3] C. North, R. Chang, A. Endert, W. Dou, R. May, B. Pike, and G. Fink, "Analytic provenance: Process+interaction+insight," in *Extended Abstracts of the 2011 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA 11. New York, NY, USA: Association for Computing Machinery, ACM, 2011, p. 3336. [Online]. Available: https://doi.org/10.1145/1979742.1979570

[4] E. D. Ragan, A. Endert, J. Sanyal, and J. Chen, "Characterizing provenance in visualization and data analysis: An organizational framework of provenance types and purposes," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 22, no. 1, pp. 31–40, 2016.

[5] K. Madanagopal, E. D. Ragan, and P. Benjamin, "Analytic provenance in practice: The role of provenance in real-world visualization and data analysis environments," *IEEE Computer Graphics and Applications*, vol. 39, no. 6, pp. 30–45, 2019.

[6] K. Xu, A. Ottley, C. Walchshofer, M. Streit, R. Chang, and J. Wenskovitch, "Survey on the analysis of user interactions and visualization provenance," *Computer Graphics Forum*, vol. 39, no. 3, pp. 757–783, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14035

[7] J. Hakanen, K. Miettinen, and K. Matković, "Task-based visual analytics for interactive multiobjective optimization," *Journal of the Operational Research Society*, pp. 1–18, 2020.

[8] H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, and B. Eiderbäck, "Technology probes: Inspiring design for and with families," in *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, CHI '03*. Association for Computing Machinery, 2003, p. 1724. [Online]. Available: https://doi.org/10.1145/642611.642616

[9] K. Xu, S. Attfield, T. Jankun-Kelly, A. Wheat, P. H. Nguyen, and N. Selvaraj, "Analytic provenance for sensemaking: A research agenda," *IEEE Computer Graphics and Applications*, vol. 35, no. 3, pp. 56–64, 2015.

[10] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala, "Graphical histories for visualization: Supporting analysis, communication, and evaluation," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 14, no. 6, pp. 1189–1196, 2008.

[11] P. H. Nguyen, K. Xu, A. Bardill, B. Salman, K. Herd, and B. L. W. Wong, "Sensemap: Supporting browser-based online sensemaking through analytic provenance," in *2016 IEEE Conference on Visual Analytics Science and Technology, VAST*, 2016, pp. 91–100.

[12] C. Dunne, N. Henry Riche, B. Lee, R. Metoyer, and G. Robertson, "Graphtrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history," in *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, CHI '12*, 2012, p. 16631672. [Online]. Available: https://doi.org/10.1145/2207676.2208293

[13] J. Zhao, M. Glueck, P. Isenberg, F. Chevalier, and A. Khan, "Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 24, no. 1, pp. 340–350, 2018.

[14] K. Xu, A. Ottley, C. Walchshofer, M. Streit, R. Chang, and J. Wenskovitch, "Survey on the analysis of user interactions and visualization provenance," *STAR*, vol. 39, no. 3, 2020.

[15] A. Skopik and C. Gutwin, "Improving revisitation in fisheye views with visit wear," in *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, CHI '05*, 2005, p. 771780. [Online]. Available: https://doi.org/10.1145/1054972.1055079

[16] M. Wattenberg and J. Kriss, "Designing for social data analysis," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 12, pp. 549–557, 2006.

[17] W. Willett, J. Heer, and M. Agrawala, "Scented widgets: Improving navigation cues with embedded visualizations," in *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, CHI '07*, 2007, pp. 51–58.

[18] M. Feng, C. Deng, E. M. Peck, and L. Harrison, "Hindsight: Encouraging exploration through direct encoding of personal interaction history," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 23, no. 1, pp. 351–360, 2017.

[19] A. Narechania, A. Coscia, E. Wall, and A. Endert, "Lumos: Increasing awareness of analytic behavior during visual data analysis," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 28, no. 1, pp. 1009–1018, 2021.

[20] E. Wall, A. Narechania, A. Coscia, J. Paden, and A. Endert, "Left, right, and gender: Exploring interaction traces to mitigate human biases," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 28, no. 1, pp. 966–975, 2021.

[21] E. Dimara, A. Bezerianos, and P. Dragicevic, "Conceptual and methodological issues in evaluating multidimensional visualizations for decision support," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 24, no. 1, pp. 749–759, 2018.

[22] J. Bautista and G. Carenini, "An empirical evaluation of interactive visualizations for preferential choice," in *Proceedings of the Working Conference on Advanced Visual Interfaces*, ser. AVI '08. New York, NY,

USA: Association for Computing Machinery, 2008, p. 207214. [Online]. Available: https://doi.org/10.1145/1385569.1385603

[23] E. Triantaphyllou, "Multi-criteria decision making methods: a comparative study," *Springer Science & Business Media*, vol. 44, 2013.

[24] B. Ceballos, M. Lamata, and . D. Pelta, "A comparative analysis of multi-criteria decision-making methods," in *Prog Artif Intell*, 2016, p. 315322.

[25] M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer, and T. Möller, "Visual parameter space analysis: A conceptual framework," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 20, no. 12, pp. 2161–2170, 2014.

[26] M. Booshehrian, T. Möller, R. M. Peterman, and T. Munzner, "Vismon: Facilitating analysis of trade-offs, uncertainty, and sensitivity in fisheries management decision making," in *Computer Graphics Forum*, vol. 31, no. 3pt3. Wiley Online Library, 2012, pp. 1235–1244.

[27] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit, "Lineup: Visual analysis of multi-attribute rankings," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 19, no. 12, pp. 2277–2286, 2013.

[28] S. Pajer, M. Streit, T. Torsney-Weir, F. Spechtenhauser, T. Möller, and H. Piringer, "Weightlifter: Visual weight space exploration for multi-criteria decision making," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 23, no. 1, pp. 611–620, 2016.

[29] T. Tuar and B. Filipi, "Visualization of pareto front approximations in evolutionary multiobjective optimization: A critical review and the prosection method," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 19, no. 2, pp. 225–245, 2015.

[30] K. Miettinen, "Survey of methods to visualize alternatives in multiple criteria decision making problems," *OR spectrum*, vol. 36, no. 1, pp. 3–37, 2014.

[31] A. Marsaw, B. German, P. Hollingsworth, D. Mavris, and S. Csonka, "An interactive visualization environment for decision making in aircraft engine preliminary design," in *45th AIAA Aerospace Sciences Meeting and Exhibit*, 2007, p. 1333.

[32] J. Sorger, T. Ortner, C. Luksch, M. Schwärzler, E. Gröller, and H. Piringer, "Litevis: integrated visualization for simulation-based decision support in lighting design," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 22, no. 1, pp. 290–299, 2015.

[33] X. Zhao, Y. Wu, W. Cui, X. Du, Y. Chen, Y. Wang, D. L. Lee, and H. Qu, "Skylens: Visual analysis of skyline on multi-dimensional data," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 24, no. 1, pp. 246–255, 2017.

[34] L. Cibulski, H. Mitterhofer, T. May, and J. Kohlhammer, "Paved: Pareto front visualization for engineering design," in *Computer Graphics Forum*, vol. 39, no. 3. Wiley Online Library, 2020, pp. 405–416.

[35] H. L. Trinkaus and T. Hanne, "knowcube: a visual and interactive support for multicriteria decision making," *Computers & Operations Research*, vol. 32, no. 5, pp. 1289–1309, 2005.

[36] A. Bezerianos, F. Chevalier, P. Dragicevic, N. Elmqvist, and J. D. Fekete, "Graphdice: A system for exploring multivariate social networks," in *Proceedings of the 12th Eurographics / IEEE - VGTC Conference on Visualization*, ser. EuroVis10. Chichester, GBR: The Eurographs Association & John Wiley & Sons, Ltd., 2010, p. 863872. [Online]. Available: https://doi.org/10.1111/j.1467-8659.2009.01687.x

[37] N. Elmqvist, P. Dragicevic, and J. Fekete, "Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 14, no. 6, pp. 1539–1148, 2008.

[38] N. Boukhelifa, W. Cancino, A. Bezerianos, and E. Lutton, "Evolutionary visual exploration: evaluation with expert users," in *Computer Graphics Forum*, vol. 32, no. 3pt1. Wiley Online Library, 2013, pp. 31–40.

[39] R. A. Rensink and G. Baldridge, "The perception of correlation in scatterplots," in *Computer Graphics Forum*, vol. 29, no. 3. Wiley Online Library, 2010, pp. 1203–1210.

[40] E. Dimara, A. Bezerianos, and P. Dragicevic, "Conceptual and methodological issues in evaluating multidimensional visualizations for decision support," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 24, no. 1, pp. 749–759, 2018.

[41] A. Barczewski, A. Bezerianos, and N. Boukhelifa, "How domain experts structure their exploratory data analysis: Towards a machine-learned storyline," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '20. Association for Computing Machinery, ACM, 2020, p. 18. [Online]. Available: https://doi.org/10.1145/3334480.3382845

[42] E. D. Ragan, J. R. Goodall, and A. Tung, "Evaluating how level of detail of visual history affects process memory," in *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, 2015, pp. 2711–2720.

[43] Y. Park, M. Cafarella, and B. Mozafari, "Visualization-aware sampling for very large databases," in *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, 2016, pp. 755–766.

[44] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[45] A. Strauss and J. Corbin, *Basics of qualitative research*. Sage publications, 1990.

[46] R. Amar, J. Eagan, and J. Stasko, "Low-level components of analytic activity in information visualization," in *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 2005, pp. 111–117.

[47] D. Gotz and M. X. Zhou, "Characterizing users' visual analytic activity for insight provenance," *Information Visualization*, vol. 8, no. 1, pp. 42–55, 2009.

[48] P.-M. Law, A. Endert, and J. Stasko, "Characterizing automated data insights," in *2020 IEEE Visualization Conference (VIS)*. IEEE, 2020, pp. 171–175.

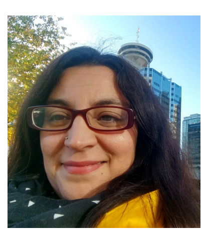
[49] Z. Liu, Y. Wang, M. Dontcheva, M. Hoffman, S. Walker, and A. Wilson, "Patterns and sequences: Interactive exploration of clickstreams to understand common visitor paths," *IEEE Transactions on Visualization and Computer Graphics, TVCG*, vol. 23, no. 1, pp. 321–330, 2016.

[50] M. Loorak, M. Tory, and S. Carpendale, "Changecatcher: Increasing inter-author awareness for visualization development," in *Computer Graphics Forum*, vol. 37, no. 3. Wiley Online Library, 2018, pp. 51–62.

[51] Y. B. Shrinivasan and J. van Wijk, "Supporting exploration awareness in information visualization," *IEEE Computer Graphics and Applications*, vol. 29, no. 5, pp. 34–43, 2009.

Mehdi Chakhchoukh is currently a PhD student with Université Paris-Saclay, Orsay, France. His work focuses on trade-off analysis and sense-making. He is the corresponding author of this article. Contact him at mehdi.chakhchoukh@universite-paris-saclay.fr.

Nadia Boukhelifa is currently a Research Scientist in visual analytics with INRAE, Paris, France. Her work focuses on uncertainty visualization, mixed-initiative systems, and visual analytics applications from agronomy and biological sciences. Contact her at nadia.boukhelifa@inrae.fr.



Anastasia Bezerianos is currently an Associate Professor with Université Paris-Saclay, Orsay, France. She works at the intersection of human-computer interaction and information visualization, with a focus on collaboration, decision making and novel displays. Contact her at anastasia.bezerianos@universite-paris-saclay.fr.