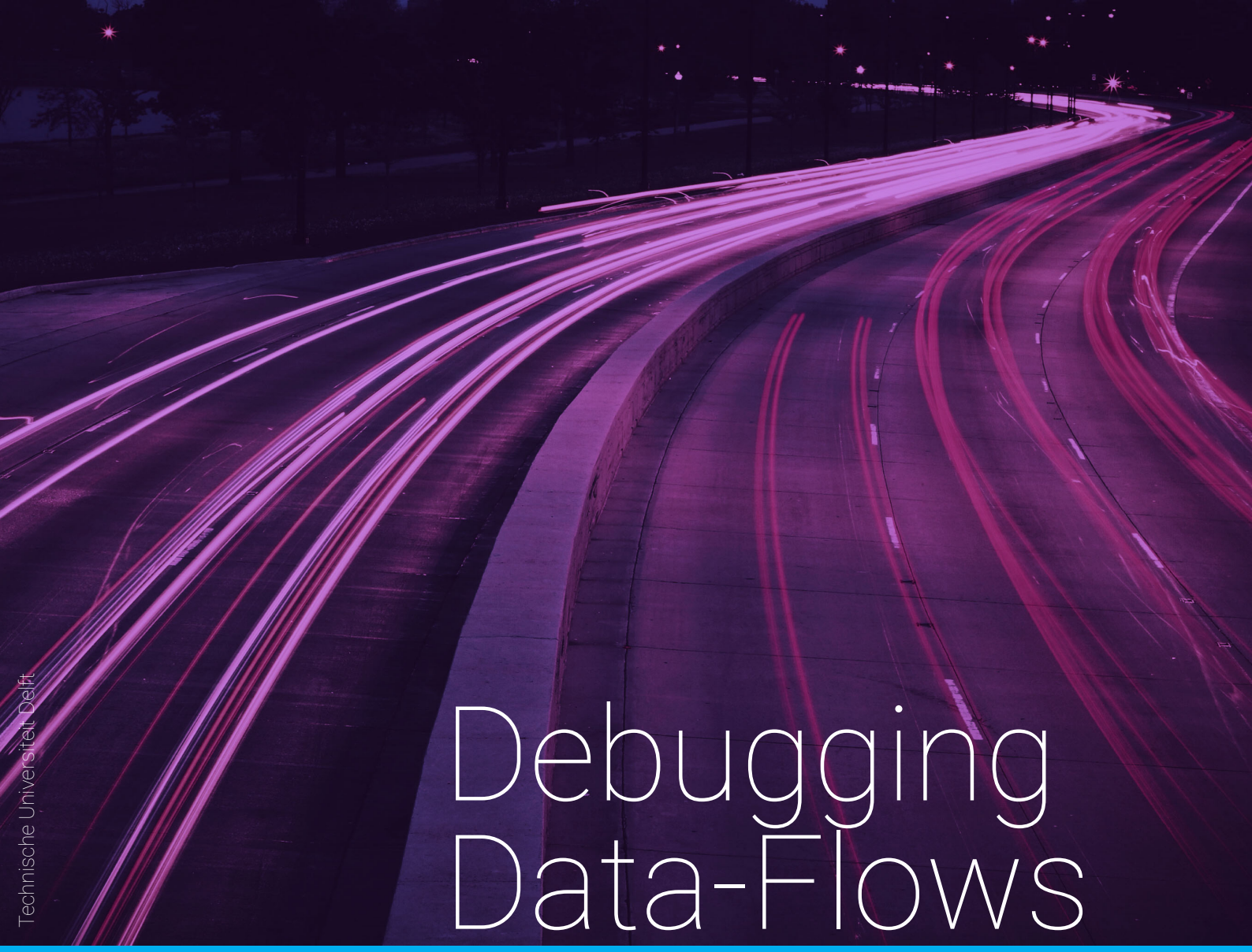


Debugging Data-Flows in Reactive Programs

Herman Banken

Technische Universiteit Delft

TU Delft

Debugging Data-Flows in Reactive Programs

by

Herman Banken

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday July 5th, 2017 at 4:00 PM.

Student number:	4078624	
Project duration:	September 12, 2016 – June 30, 2017	
Thesis committee:	Prof. Dr. H.J.M. Meijer,	TU Delft
	Dr. G. Gousios,	TU Delft
	Prof. Dr. A van Deursen,	TU Delft
	Joost de Vries,	Ordina

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft Delft University of Technology

CODE.STAR

Acknowledgements

I would like to thank everyone involved in my thesis project for their support. First of all, Erik Meijer, thank you for the opportunity to do a thesis about Rx, inspiring me to use the *hacker mentality* to create RxFiddle, and for supporting me throughout this project, from across the Atlantic.

I want to thank Georgios Gousios for his belief in me, the opportunity to write a paper instead of a report, and for his time, guidance, support and detailed feedback. Thanks to Arie van Deursen for taking the time to be part of my thesis committee.

Furthermore I want to thank everyone at Codestar, in particular Joost de Vries, Martin Weidner and David Hoepelman, for their time and support, the great discussions at Fridayś in Nieuwegein, and making this research possible. Joost and David, thank you for getting me up to speed and keeping me on track, especially in those first months when exploring the many possibilities ahead.

Thanks to Mark, Leonard, Rik, Joseph, Richard, Mike, Georgi and Eddy, and everyone else who provided me with invaluable feedback during this project.

Finally, I would like to thank my family and friends for their support during my studies and my thesis.

Herman Banken
Delft, June 2017

Debugging Data-Flows in Reactive Programs

Herman Banken
Delft University of Technology,
The Netherlands
hermanb@ch.tudelft.nl

Georgios Gousios
Delft University of Technology,
The Netherlands
g.gousios@tudelft.nl

Erik Meijer
Delft University of Technology,
The Netherlands
h.j.m.meijer@tudelft.nl

ABSTRACT

Reactive Programming is a way of programming designed to provide developers with the right abstractions for creating systems that use streams of data. Traditional debug tools lack support for the abstractions provided, causing developers to fallback to the most rudimentary debug tool available: `printf`-debugging.

In this work, we design a visualization and debugging tool for Reactive Programming, that aids comprehension and debugging of reactive systems, by visualizing the dependencies and structure of the data flow, and the data inside the flow. We present RxFiddle, a platform for the visualization as well as the required instrumentation for RxJS in the ReactiveX-family of Reactive Programming libraries. Evaluation based on an experiment with 111 subjects, shows that RxFiddle can outperform traditional debugging in terms of debug time required.

Keywords

reactive programming, debugging, visualization, program comprehension

1. INTRODUCTION

Software often needs to respond to external events and data flows. Consider for example software in interactive applications, for desktops, web and mobile phones, in graphics and in processing sensor data from phones or IoT-devices. Traditionally, handling these asynchronous events was done using the *Observer design pattern* [25] or *callback functions* [19]. Using these patterns, the system consuming the data does not have to block waiting for methods to return, but instead receives a notification event when data is available. While these patterns decouple the consumer from the producer of the data, they typically lead to dynamic registration, side effects in the consumers, and inversion of control [14, 42].

Reactive Programming (RP) is an alternative to these patterns for event driven computation. RP defines event streams as lazy collections and provides operators that allow developers to deal with the complications of asynchronous event handling. It offers declarative and concise syntax for composing streams of data, to express the complex reactive behavior of these applications. RP started in academia in the form of Functional Reactive Programming (FRP) [12, 15, 16, 32, 35], but in recent years the use of RP has exploded. Languages such as Elm [11] and libraries such as Reactor [23], Akka [6] and Rx [34] are being used by companies like Netflix, Microsoft and Google, to build highly responsive and scalable systems. Front-end libraries like Angular[1], that use RP in their foundations, are used by many large sites (9.1% of Quantcast Top 10k websites[2]). Developers and companies alike attempted to standardize "Reactive Programming" in the form of the Reactive Manifesto [5].

While reactive programs might be more declarative and concise, RP does not work well with traditional interactive debuggers, shipped with most IDEś [44]. RP borrows from Functional Programming (FP) for its abstractions, its laziness and advocating the use of "pure" lambda functions. Those features contribute to a control flow that is hidden inside the RP implementation library and lead to non-linear execution of user code. This results in not useful stack traces, while breakpoints do not help either, since relevant variables are frequently not in scope. Furthermore, using a low level debugger makes it harder to interact with the high level abstractions that RP provides. Compared to imperative programming, there is limited scientific knowledge regarding how developers debug reactive programs. Traditional imperative program debugging practices [2] do not apply to RP [44].

In this work we address the issue of RP debugging by designing and implementing a high level debugger called RxFiddle for a popular version of RP, namely Reactive Extensions (Rx). RxFiddle (1) provides an overview of the dependencies in the data flow, (2) enables a detailed insight in the data flow and the timing of individual events, and (3) allows developers to trace values back through the data flow. To guide our design we conducted interviews among professional developers. After building RxFiddle, we validated it with a user experiment. The results show that RxFiddle can help developers debug RP data flows faster.

[1] `https://angular.io/`
[2] `https://trends.builtwith.com/`, accessed 2017-06-20

To steer the research, we formulate the following research questions:

RQ1 How do developers debug RP?
Before we design tools it is important to understand the problems arising in the the current state [46]. Anecdotal evidence by a number of resources[3][4] suggests that debugging RP is difficult.

RQ2 How can we design a tool that helps developers debug RP?
By examining the results of RQ1, the limitations of traditional debuggers and the opportunities that RP programs offer in terms of structure and explicit dependencies between data flows, we design a novel RP debugger.

RQ3 Can specialized RP debuggers speed up comprehension & debugging?
To validate our design and examine whether specialized tooling can improve the experience we measure the speed and correctness of comprehension in an experiment.

2. BACKGROUND: RP

Reactive Programming (RP) is a declarative programming paradigm for working with streams of input data. According to the first definition[5] a reactive program must interact with the environment "at a speed which is determined by the environment". Conceptually, when a reactive program is run, it sets up a data pipeline and waits until input arrives, i.e. when the environment changes. Reactive Programming languages and libraries provide developers with a set of abstractions and methods to create such programs.

The programming paradigm of Reactive Programming is implemented by multiple languages and libraries. Many RP implementations share a notion of a collection that abstracts over *time*, in contrast to *space* like standard collections. This collection comes in different flavors, such as Observable (Rx [34]), Signal (Elm [11]), Signal/Event (REScala [43]) or Behavior/Event (FRP [16]). The implementations differ in the precise semantics of their collections, their execution model (push/pull), and the set of available operators. In this paper, we focus on the Rx formulation, but our work is applicable to other RP implementations to some extent.

Understanding how we derive our visualization requires a minimal understanding of Rx. Rx introduces two basic types, *Observable* and *Observer*. Observables define the data flow and produce the data while Observers receive the data, possibly moving the data further down the stream. Figure 1a shows a very basic example of an "in situ" data flow in Rx. Initially, an Observable is created, here using the static `of`-method, then dependent Observables are created using the `map` and `filter`-methods on the Observable instance. Finally we `subscribe` to start the data flow and send the data in the flow to the console.

[3] `http://contributordays.com/contributor-days/rxjs`
[4] `https://staltz.com/how-to-debug-rxjs-code.html`
[5] "Reactive programs [..] maintain a continuous interaction with their environment, at a speed which is determined by the environment, not the program itself." [3]

Assembly. It is important to note that Observables are lazy; initially they only specify the blueprint of a data flow. Creating this specification is called the *assembly* phase. In the code sample of Figure 1a the assembly phase consists of the calls to `of`, `map` and `filter`, creating respectively Observables o_1, o_2 and o_3 from Figure 1b.

Subscription. When the `subscribe`-method of an Observable is called, the data flow is prepared by recursively subscribing "up" the stream: every subscribe call creates an *Observer*, that is passed to the input Observable, which again subscribes an Observer to its input Observable, until finally the root Observables are subscribed to. We call this the *subscription* phase. In Figure 1a, inside the single `subscribe` call, Observer s_1 from Figure 1b is created, and passed to o_3, which in turn will recursively subscribe to o_2 with a new Observer s_2 with destination s_1, until the full chain is subscribed.

Runtime. After the root Observables are subscribed to, they can start emitting data. This is the *runtime* phase. Depending on the nature of the Observable this might attach event listeners to UI elements, open network connections or start iterating over a list of elements. Events are pushed to s_3, to s_2 and finally to s_1 which calls `console.log` in Figure 1a.

Rx identifies three types of events that can occur during the runtime phase: *next*-, *error*- and *complete*-events. Next-events contain the next value in the flow, an error-event encapsulates an error and is a unsuccessful termination to a stream, while a complete-event denotes the successful termination of the stream. There are restrictions on their order: a Observable may first emit an unlimited amount of next-events, and then either an error or a complete event. Observables do not need to emit any next-events, and do not need to terminate.

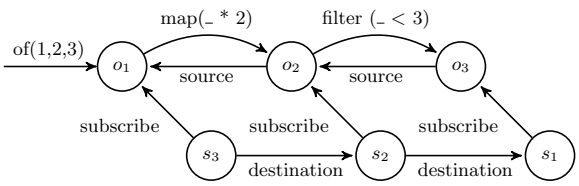
More complex programs feature operators that merge Observables[6], split Observables[7] or handle higher order Observables[8], resulting in more complex graphs. An example of `flatMap` is shown in Figure 1c. While merging and splitting happens on an Observable level (the `source` property still points to one or more dependencies), higher order Observable flattening manifests only in Observer structure (there is no reference between the Observables). Figure 1d shows this with an (abbreviated) `inner` Observable that is subscribed twice (for both values 2 and 3, value 1 is skipped), resulting in two identical data flows over o_1. The data flow through $s_{4,n}$ and s_{4_m} is pushed into s_1, flattening the data flow.

Marble Diagram. The term *Marble Diagram* comes from the shape of the glyps in the images used to explain Rx in the official documentation. An example is shown in Figure 2. The diagrams contain one or more timelines containing the events that enter and leave Observables. Next-events are typically represented with a circle, an error-event with a

[6] `concat`, `merge`, `combineLatest`, `zip`
[7] `partition`, or through multicasting with `share` or `publish`
[8] `flatMap`, `mergeMap`, `concatMap`

```
Observable.of(1, 2, 3)
  .map(x => x * 2)
  .filter(x => x < 3)
  .subscribe(console.log)
```

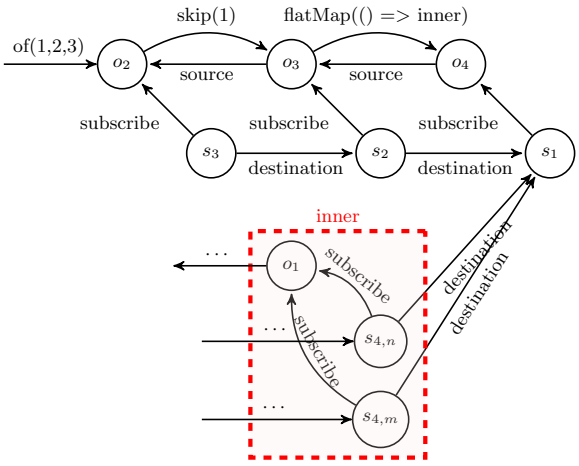
(a) Rx code example



(b) Rx graph example

```
let inner = Rx.Observable.of("A", "B")
let outer = Rx.Observable.of(1, 2, 3)
  .skip(1)
  .flatMap(() => inner)
  .subscribe()
```

(c) Higher order flatMap operation



(d) Higher order Rx graph example

Figure 1: Samples of Rx Observables

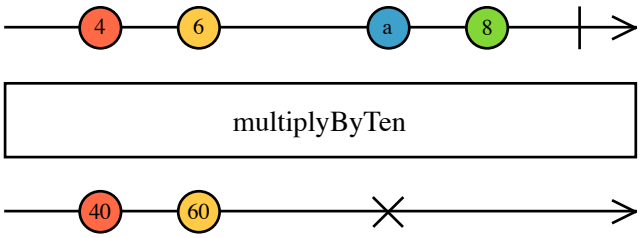


Figure 2: Marble Diagram

	Method	Focus
RQ1	Interview	What are current practices
	Literature	What is recommended
RQ2	Design	What can a RP debugger show
	Implement	Extract meta information from Rx
RQ3	Experiment	Quantification of effect on debugging

Table 1: Research Methods used in the study

cross and a complete-event with a vertical line. Developers can see from the diagram how operators work by inspecting the difference between the timelines, where events might be skipped, added, transformed or delayed. Mapping time on the x-axis provides insight that is missing when inspecting only a single time slice.

3. RESEARCH DESIGN

To answer our research questions, we employ a three-phase Sequential Exploratory Strategy, one of the mixed methods research approaches [10, 24], as shown in Table 1. First, we interview professional developers and review available documentation (RQ1) to form a understanding about current debugging practices. Second, we apply this understanding to design a debugger and implement it to test its feasibility (RQ2). Finally, we validate the debugger using an experiment (RQ3).

4. RP DEBUGGING PRACTICES

To validate the need for better tools we must first understand how existing tools are used (RQ1). We interview developers, as we want to explore and understand the current practices, instead of using an experiment or survey to test a particular hypothesis. The questions were semi-structured. We first established a general understanding of the experience of the subjects. Then we asked several open questions regarding use of RP, how subjects debug RP and test RP. Table2 lists the questions used as a guideline for the interviews.

Five developers with professional programming experience ranging from 4 to 12 years were interviewed. The first four developers (D1-D4) work in Company A, which builds mostly reactive systems [5] using various RP solutions. Developers range from a month to over a year of Rx experience. The fifth developer (D5) works in Company B, and is concerned with building and maintaining a large scale distributed server application, that uses Rx to handle asynchronous events.

4.1 Interviews

In the following paragraphs we discuss the results of Q6-Q10 in detail. Not every subject answered each question in the same amount of detail, so we discuss the answers that provide meaningful insights in the current practice.

	Question	
Q1	Explain your (professional) experience.	
Q2	Assess your experience on a scale from beginner to expert.	Context,
Q3	Explain your (professional) reactive programming experience.	understanding
Q4	Assess your RP experience on a scale from beginner to expert.	subjects
Q5	Did you refactor or rework RP code?	
Q6	Did you and how did you test or verify the workings of RP code?	
Q7	Did you and how did you debug RP code?	
Q8	Did you and how did you use documentation on RP?	Content questions
Q9	What difficulties did you experience with RP?	
Q10	What is your general approach to understand a piece of Rx?	

Table 2: Interview questions

Testing. Of the 4 subjects of Company A, none performed tests specifically for Rx logic. *"Just running the application"*, is enough according to D3, saying that they only test the business logic in their application and consider the Rx code as "glue"' which either works or does not work. In contrast, D5 and his team at Company B extensively tests their application using the Rx' built-in test facilities like "marble tests" and the `TestScheduler` [40]. Using tests, the subject *confirms his believes about the behavior* of the chain of operators, and tests also help later on when refactoring code.

Debugging. All subjects independently mention using temporary `printf`-debugging statements (`console.log` in JavaScript). Subjects use `printf`-debugging to *"add more context"* (D1) to their debug sessions. Printing which values flow through the flow allows them to *"quickly reason what happens"* (D3). Breakpoints are only used when the project requires costly (TypeScript) recompilation if the source is modified, as required for `printf`-debugging (D1).

Existing debuggers often can not be used to inspect the life-cycle of Observables (subscription and disposal), as these occurrences are not normally defined in user code and would require breakpoints in library code, like the `subscribe`-method, which is used by all class instances of Observable. This debugging inside the Rx library was described as *"painful"* by D2, when using the Node.js debugger to step through the inners of Rx. Alternatives used by our subjects are (1) creating a *custom `debug`-operator* which prints these life-cycle events (D5) or (2) creating custom Observables (`Observable.create`) with *custom `subscribe`- and `dispose`-methods*, inserted at the beginning of the chain, that print upon their usage (D2, D5). While `printf`-debugging and breakpoints are useful in various degrees when executing a single Observable chain, these methods both become considerably more difficult and *"overview is easily lost"* when executing multiple chains concurrently (D3, D5).

Documentation. Subjects give different reasons to visit the documentation, but the most common reason is to *"find an operator for what I need"* (D1). They feel that there might be an operator that precisely matches their needs, however knowing all operators by heart is not common (the Rx's Observable API has 28 static methods and 114 instance methods), therefore subjects sometimes end up doing an extensive search for some specific operator. Another reason to visit the documentation is to *comprehend how operators in existing code work*. For this, subjects use the Marble Diagrams at RxMarbles.com [33] (D2, D5), the RxJS 4 documentation at GitHub (D2, D5), the RxJS 5 documentation at ReactiveX.io [40] (D1, D4, D5) and the online book IntroToRx.com [7] (D4). D1 specifically mentions the need for more examples in the documentation.

Difficulties experienced. The IDE does not help with developing Rx (D2, D4); according to D4 *"Rx is more about timing than about types"*, and *"You miss some sort of indication that the output is what you expect"*. It is not always clear what happens when you execute a piece of code, "mostly due to Observables sometimes being lazy" (D2). Flows are clear and comprehensible in the scope of a single class or function, but for applicationwide flows it becomes unclear (D3, D4 and D5). D3 mostly used RxScala and mentions that creating micro services helps in this regard. D1 mentions that *"you need to know a lot as a starting [RxJS] developer"*, giving the example of the many ways to cleanup and `unsubscribe`, which he did manually initially. D1 used both logging while analyzing existing code and learning to overcome inexperience.

Understanding. Subjects first *look at which operators are used*, then they *reason about what types and values might flow through the stream* (D2, D3, D4 and D5), using various methods. By analyzing the variable names D2 forms an expectation of the resulting value types, then reasoning backwards, to see how this data is derived. *Running the code*, is used when possible by D5, to observe the outcome of the stream, as this "shows the intentions of the original developer". If it remains unclear how the data is transformed, the subject adds his `debug`-operator or looks up operators in the documentation.

4.2 Analysis of Literature

Developers can learn Rx through several sources such as the official documentation at ReactiveX.io, books, online courses and from the many blog posts available. We gathered resources to be analyzed by selecting 4 popular books about Rx, and complement this with the official documentations and an article by a core contributor of RxJS. All reviewed resources either mention debugging briefly and suggest using the `do`-operator for `printf`-debugging, or teach the developer `printf`-debugging via code samples.

The RxJS 4 documentation[9] and two books [17, 38] propose the use of the `do`-operator for debugging. Esposito and Ciceri [17] further explain how to best format the log statements and introduce ways to limit the logging by modifying the Observable through means of throttling and sampling. The RxJava book [38] also contains tips to use the various `do`-operators to integrate with existing metric tools. To our knowledge the only article[10] addressing issues of debugging Rx is by Staltz, one of the contributors of RxJS, noting that conventional debuggers are not suitable for the higher level of abstraction of Observables. Staltz explains three current ways to debug Rx, being: (1) tracing to the console, (2) manually drawing the dependency graph, (3) manually drawing Marble Diagrams.

We analyzed a set of 13 books about RxJS, which was created by selecting 69 books matching "RxJS" from the O'Reilly Safari catalogue[11], and further reducing the set by filtering on the terms "debug" and "debugger". While, none of the remaining books had a chapter about debugging, many of these books use `printf`-debugging in their code samples. Notably, Blackheath [4] suggests, in a "Future Directions" chapter, that special debuggers could provide a graphical representation of FRP state over time and would allow debugging without stepping into the FRP engine.

4.3 Overview of practices

The available literature matches the results of the interviews: `printf`-debugging is commonly advised and used. While the conventional debugger works in some cases, this is mostly the case for the procedural logic that interleaves Rx logic. Automated tooling is suggested, but is not implemented. We see that developers use `printf`-debugging to learn the behavior of Observables, behavior meaning both their values flowing through and their (one or many) subscriptions.

Overall, we identified four overarching practices when debugging Rx code:

(1) Gaining high-level overview of the reactive structure.

(2) Understanding dependencies between Observables.

(3) Finding bugs and issues in reactive behavior.

(4) Comprehending behavior of operators in existing code.

5. DEBUGGER DESIGN

In this section we describe the design of a visualizer for the ReactiveX (Rx) family of RP libraries to answer RQ2. Given the findings of RQ1, the requirements for our visualizer are:

REQ1 Provide overview of Observable flows. This overview should support practices 1 and 2, by graphically representing the relations between Observables, such that a complete image is given of all Observables and how they interact.

REQ2 Provide detailed view inside flow. This view should support practices 3 and 4 by giving access to both data flow and life-cycle events and should be able to show the behavior of a operator visually.

To meet those requirements, we propose a visualizer consisting of two parts: (1) a Data Flow Graph and (2) a Dynamic Marble Diagram. The data flow graph satisfies REQ1, providing a high-level overview, showing how different flows are created, combined and used, while the marble diagram satisfies REQ2, offering a more in-depth look into a single selected data flow showing the contents (in terms of values and subscriptions) of the flows and can be used to learn the behaviors and interplay of operators.

5.1 Data Flow Graph

Simplified graphs. When running a RP program, Observables are created that depend on other Observables (their *source*) and Observers are created to send their values to a defined set of Observers (their *destination*). Figure 1b shows these relations in a graph. For the simplest of programs, the relations between the Observables ($O = o_1, o_2, o_3$) and those between Observers ($S = s_1, s_2, s_3$) share an equally shaped sub-graph after a reversal of the Observer-edges. To provide more overview, we process the graph to merge the two Observable and Observer sequences together, simplifying it in the process, resulting in a *Data Flow Graph* (DFG) as in Figure 3. The process is simple: we retain only the Observer-subgraph nodes, complementing them with the meta data of the corresponding Observable nodes. Higher order relations are retained, as shown in Figure 4. Figure 5B shows the DFG in practice.

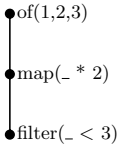


Figure 3: DFG of Figure 1b

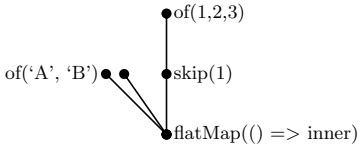


Figure 4: DFG of Figure 1d

Layout. Layout is used to add extra meaning to the graph. If multiple subscriptions on the same Observable are created, multiple flows are kept in the graph and they are bundled together in the resulting layout. This is designed to help developers find related flows. Also it is easy to see that for example an Observable is reused many times, hinting a possible performance improvement by sharing the computation (Rx has special `share`-operators to multicast). The layout is based on StoryFlow [30], which employs a hierarchical clustering before ordering the graph in a way to reduce

[9] `https://github.com/Reactive-Extensions/RxJS/blob/master/doc/gettingstarted/testing.md`
[10] `http://staltz.com/how-to-debug-rxjs-code.html`
[11] `http://www.safaribooksonline.com`

crossings. Whereas StoryFlow clusters on physical character location, we cluster flows per Observable. Furthermore, StoryFlow supports interactivity in various layout stages of which we use the algorithms for *straightening* and *dragging* to support selecting a specific flow, which is then highlighted, straightened and positioned at the right in order to match the Marble Diagram, shown for the current highlighted flow.

Color. Coloring the nodes can be used to identify the same Observable in multiple places in the graph, as Observables can be reused in different places of the stream. For example, in Figure 1d the `inner` Observable is reused twice, which we denote visually by applying the same color to its two occurrences in the DFG.

5.2 Dynamic Marble Diagrams

In contrast to the original diagrams (Section2), we use dynamic diagrams which update live when new events occur and are stacked to show the data in the complete flow. This allows the developer to *trace a value back through the flow*, a debug operation which is impossible using a classic debugger. Handcrafted marble diagrams can use custom shapes and colors to represent events, but for the generic debugger we use only three shapes: next-events are a green dot, errors are a black cross and completes are a vertical line, as shown in Figure 5C. For our generic debugger, it is unfeasible to automatically decide which properties (content, shape and color) to apply to events, as the amount of events and distinguishing features might be unbounded. Instead the event values are shown upon hovering.

5.3 Architecture

To support the visualization, we design a debugger architecture consisting of two components: a host instrumentation and a visualizer.

The **Host instrumentation** instruments the Rx library to emit useful execution events. Depending on the language and platform, specific instrumentation is required. Output of the instrumentation is a platform and language independent graph like Figure 1d. By splitting the instrumentation from the visualization, the debugger can be used for the complete Rx family of libraries by only reimplementing the first component. The communication protocol for the instrumentation is shown in Table 3. Note that the user never needs to use this protocol, it is internal to the debugger.

The **Visualizer** takes the output of the host instrumentation, the initial graph, and simplifies it into a Data Flow Graph. Then it lays out the Data Flow Graph and provides the debuggers User Interface. By separating the visualizer, we can safely export generated graphs and visualize them post mortem for example for documentation purposes.

The components can run in their own environment. The instrumentation must run inside the host language, while the Visualizer can use a different language and platform.

5.4 Implementation

To validate the design and to provide an implementation to the developer community we created `RxFiddle.net`. The RxFiddle project is a reference implementation of our reactive debugger design. Besides the visualizer, the website also contains a code editor for JavaScript code with sharing functionality, for developers to share snippets with their peers, as shown in Figure 5A. In this section we will explain different parts of the implementation. For RxFiddle, we initially focused on RxJS (JavaScript).

Instrumentation. With JavaScript being a dynamic language, we use a combination of prototype patching and Proxies[12] to instrument the RxJS library: the Observable and Observer prototypes are patched to return Proxies wrapping the API method calls. The instrumentation passes every method entry and method exit to the Linking-step.

Linking. We distinguish between method calls from the different phases (Section 2). From the assembly phase, we detect when Observables are used as target or arguments of a call or as return value, and create a graph node for each detected Observable. We add an edge between the call target & call arguments and returned Observables, denoting the *source*-relation. Also, we tag the returned Observable with the call frame information (time, method name, arguments). In the subscription phase, we detect calls to the `subscribe`-method: the destination Observers are passed as arguments, so we create the graph nodes and save the relation as an edge. In the runtime phase we detect `next-`, `error-` and `complete`-calls on Observers and add these as meta data to the Observer nodes.

Graph Loggers. From the Linking-step the graph mutations are streamed to the environment of the visualizer, where the graph is rebuilt. Depending on the host language a different protocol is used: RxFiddle's code editor executes the code in a Worker[13] and transmits events over the postMessage protocol, while RxFiddle for Node.js transmits over WebSockets. Being able to support multiple protocols, extends the possible use cases, ranging from the code editor for small programs, to a Node.js plugin for server applications, to Chrome DevTool extensions[14] for web applications.

Visualizer. The visualizer receives the current state in the form of a graph from the Logger. It then uses the Observers in the graph to create the DFG. To layout the DFG using StoryFlow [30], we first rank the graph using depth first search, remove slack [20] and reverse edges where necessary to create a directed acyclic graph. We then add dummy nodes to replace long edges with edges spanning a single rank. Finally we order and align the nodes in the ranks assigning coordinates for the visualization. It is important that layouting is fast, as it runs every time the DFG is changed. To render the Marble Diagrams, the flow *to* and *from* the selected Observer is gathered, by recursively traversing the graph in the direction of the edges, respectively the reversed direction.

[12] `https://developer.mozilla.org/docs/Web/JavaScript/Reference/Global_Objects/Proxy`
[13] `https://developer.mozilla.org/docs/Web/API/Worker`
[14] `https://developer.chrome.com/extensions/devtools`

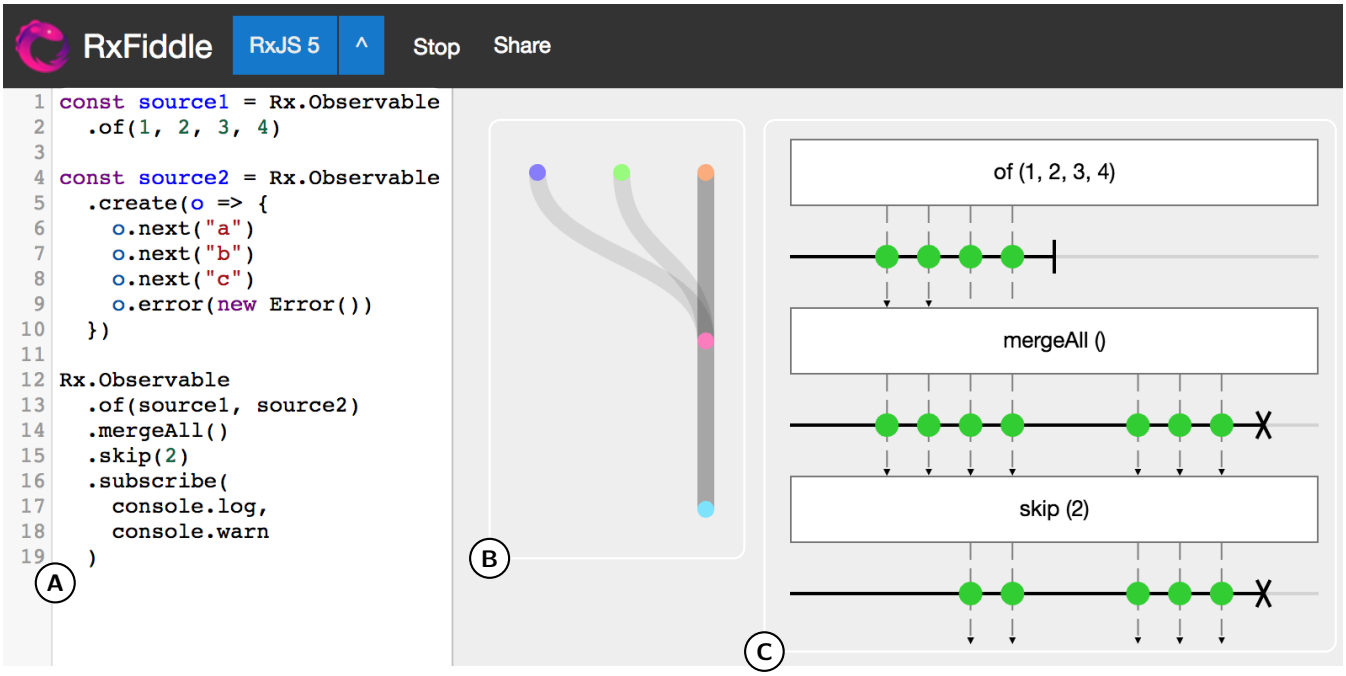


Figure 5: Screenshot of RxFiddle.net, showing the Code Editor (A), the DFG (B) and the Dynamic Marble Diagram (C)

`addObservable(id, sourceIds)`	Adds a Observable node, with zero or more source Observable's
`addObserver(id, observableId, destinationId)`	Add a Observer, `observableId` denotes the Observable it subscribed to, optional `destinationId` adds an edge to the destination Observer
`addOuterObserver(observerId, outerDestination)`	Create a special edge between an existing Observer and the higher order destination Observer
`addEvent(observerId, type, optionalValue)`	Add an event to the Observer denoted by `observerId`, of type (next, error, complete), optionally with a value (for next / error events).
`addMeta(id, metadata)`	Add meta data such as the method call which created an Observable.

Table 3: Instrumentation protocol

6. EVALUATION

In this section we evaluate our debugger to assess the efficacy of our approach, by answering RQ3. We want to compare the original situation with the situation when using RxFiddle. RxFiddle is our "treatment" for the RP debugging problem, so we use an experiment, in which we control for the debugger that subjects use, to test our hypothesis that RxFiddle improves debugging.

Ko et al. [28] describes two commonly used measures for experiments regarding tools in Software Engineering: *success* on task, and *time* on task. The goal of our experiment is to measure the *time* required to solve programming problems correctly. If our reasoning for RQ2 is right and our design leans itself for RP, we expect to see that the group using RxFiddle can more quickly reason about the reactive code at hand and can trace bugs faster. We do not choose only success or correctness as a measure for the experiment, as we expect both groups to be able to complete the tasks correctly: while the current debugging situation is non-optimal, it is still used in practice, indicating that it works at least to some extend. The construct of time also matches debugging better; a developer needs to continue debugging *until* he finds an explanation or a solution to his problem, and incorrect assumptions can be tested and corrected.

We measure the time from the moment the participant receives the question until the correct answer is given. Participants use either the built-in Chrome Browser debugger (group *Console*) or - the treatment - our RxFiddle debugger (group *RxFiddle*). This single alternative Console debugger together with the experiment UI (which acts as a small IDE) offers all the debugging capabilities subjects of our preliminary interviews (RQ1) reported to use.

The experiment consists of a questionnaire, a warm-up task and four programming tasks, all available in a single in-browser application, of which the source code is available at [1]. The questionnaire contains questions regarding age, experience in several programming languages and several reactive programming frameworks. We use this self estimation as a measurement of skill instead of a pretest, since it is a faster and better estimator [18, 27, 45]. The warm-up program is situated in the same environment as the programming problems and contains several tasks designed to let the participants use every control of this test environment. The first two programming problems require the participants to obtain an understanding about the behavior of the program and report the findings. The last two programming problems contain a program with a bug. The participants are asked to find the event that leads to the bug in the third

problem and to identify and textually propose a solution in the fourth problem. The first two problems are synthetic examples of two simple data flows, while the latter two contain some mocked (otherwise remote) service which behaves like a real world example.

We use a between-subjects design for our setup. While this complicates the results - subjects have different experience and skills - we can not use a within-subjects design as it would be impossible to control for the learning effect incurred when asking subjects to perform survey questions with and without the tool. This also allows us to restrict the amount of tasks to incorporate in the experiment, requiring less time from our busy subjects. In the experiment environment subjects can answer the question and then hit "Submit"; alternatively they can "Pass" if they do not know the answer.

6.1 Context

The experiment was run both in a offline and in an online setting. The offline experiment was conducted at a Dutch software engineering company. Subjects are developers with several years of programming experience, and range from little to no experience with RP to many years of experience (Figure 6). As we do not try to measure the effect of learning a new tool, but rather using a tool after learning to use it, we explained RxFiddle in the introductory talk and added the warm-up question to get every participant to a minimum amount of knowledge about the debugger at hand.

The online experiment was announced by the authors on Twitter, and consequently retweeted by several core contributors to RP libraries, and via various other communication channels, such as Rx related Slack and Gitter topics. Subjects to the online experiment took the test at their own preferred location and have possibly very different backgrounds. We created several short video tutorials and included these in the online experiment to introduce the participants to the debug tool available to them and the tasks they needed to fulfill. The introductory talk was used as the script for the videos, in order to get all participants to the same minimum level of understanding.

6.2 Results

The online experiment was performed outside of our control, and some participants quit the experiment prematurely. In total we had 111 subjects (13 offline, 98 online) starting the survey, of those 98 completed the preliminary questionnaire, and 89, 74, 67, and 58 subjects started respectively T1, T2, T3 and T4. All of the subjects in the offline setting started all tasks. Figure 7 shows the outcome of the tasks; in the remainder of this section consider only the outcomes marked as "Correct".

Overall. Figure 8 shows the time until the correct answer was given per task. Here we consider the combined results from the offline experiment and the online experiment. We make no assumptions about the underlying distribution so we perform a non-parametric Wilcoxon Mann-Whitney U test (*H_0: times for the Console group and RxFiddle group are drawn from the same population*) to see if the differences are significant, and a Cliff's delta test for ordinal data to determine the effect size. The results are shown in Table 4.

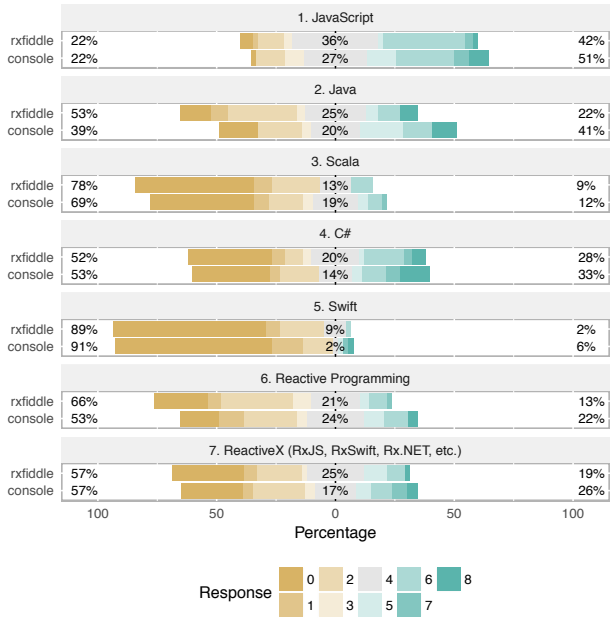


Figure 6: Experience in various programming languages, 9-point Likert scale (0 = none, 2 = beginner, 4 = medium, 6 = senior, 8 = expert)

For tasks T3 we can reject H_0 with high significance ($p < 0.05$), *the RxFiddle group is faster.* For the tasks T1, T2 and T4 we can not reject H_0 ($p > 0.05$), meaning the *RxFiddle group and Console group perform or could perform equally.*

Control for experience. To investigate this further, we split the results for different groups of subjects, as shown in Figure 10 in the Appendix. When we control for the self-assessed Rx experience, we see bigger differences for all tasks for groups with more experience, as shown in Figure 9 and Table 5 (we split at the median; `exp_rx` > "Beginner"-level).

	n_1	n_2	W	p-value	Cliff's δ
T1	34	36	559	0.540	0.0866
T2	32	31	517	0.780	-0.0424
T3	23	28	96	$6.19e^{-6}$	0.702
T4	13	12	60	0.347	0.231

Table 4: Results comparing the Console and RxFiddle groups, with respectively n_1 and n_2 subjects.

	n_1	n_2	W	p-value	Cliff's δ
T1	16	17	105	0.276	0.228
T2	14	13	99	0.720	-0.0879
T3	10	11	10	$7.88e^{-4}$	0.818
T4	8	7	13	$9.34e^{-2}$	0.536

Table 5: Results comparing the Console and RxFiddle groups, with respectively n_1 and n_2 subjects, with Rx experience above "Beginner"-level.

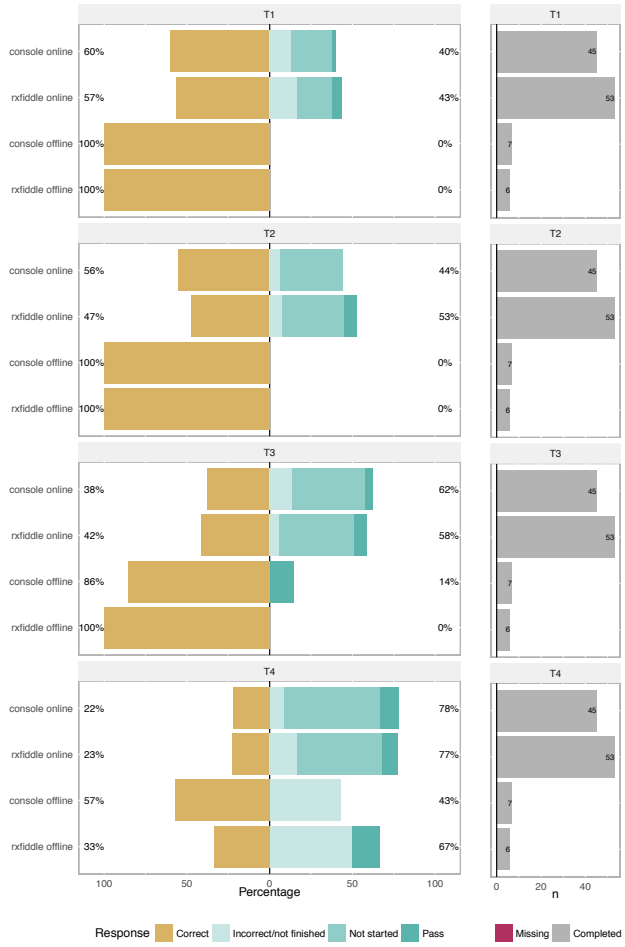


Figure 7: Task outcome per group (RxFiddle/Console) and environment (online/offline)

Still, for tasks T1, T2, and T4 we can not reject H_0, but the results are more significant comparing only experienced subjects.

7. DISCUSSION

We now discuss our main findings, how RxFiddle resolves the debugging problem of Rx, and contrast our design to other design choices and possibilities of future work.

7.1 Main results

Quick and dirty debugging. Through interviews and literature we establish that current debugging practices for RP consist mostly of `printf`-debugging. The shortcomings of this method were evident from the interviews: it works reliably only for synchronous execution or small amounts of events being logged, otherwise overview is lost. Furthermore the time-context of events and dependency-context of flows are not available using this method. We attribute the prevalence of `printf`-debugging to this *"quick and dirty"* method being available in every language and on every platform, without a viable alternative.

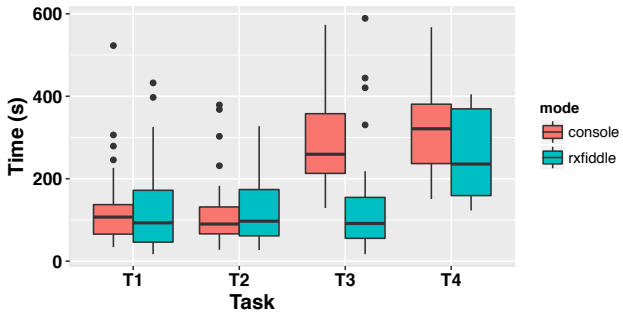


Figure 8: Time until correct answer per task, overall

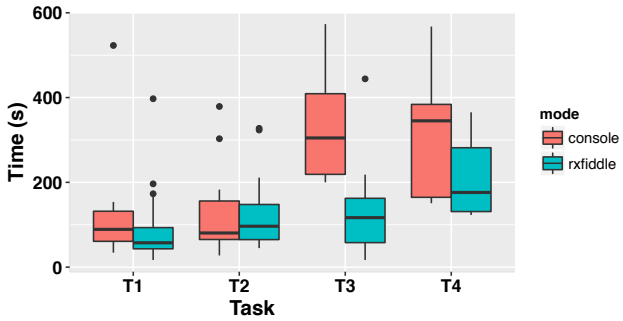


Figure 9: Time until correct answer per task, for subjects with more than "Beginner"-level of experience with Rx.

Improved context: being complete, disposing doubts. With our design and complementary implementation we show that the abstract model of RP is suitable for visualization on two levels: overview and detail. On the overview level, we complement the dependencies visible in source code with a graph of the resulting structure, showing the run-time effect of certain operators on the reactive structure. On the detail level we add the time-context, by showing previous values on a horizontal time line, and the dependency-context, by showing input and output flows above and below the flow of interest. While the results of our evaluation could be observed as a negative, RxFiddle is a new tool, where subjects have only just been exposed to the tool and received only a short training. We expect that by designing a debugger model so close to the actual abstractions, our debugger works especially well for users with some knowledge of these abstractions; while only T3 shows better performance with high significancy, we observe slightly better results when controlling for experience. Future research might investigate the effect of experience in more detail, including the use of more complicated tasks, with larger samples.

In the presented research, we did not perform tests with subjects using their own code. However, during piloting and after the release of RxFiddle we received positive feedback regarding the completeness of the visualization. As one user put it, *"by using RxFiddle when learning and understanding what RxJS does in our project, I have a feeling of improved control over our Observables, Subscriptions and the reactive parts of our app"*. Specifically the life-cycle events, which are generally hard to debug using `printf`-debugging, are more clear: *"Initially we were reluctant to manually subscribe, but*

after seeing that 'complete' often triggers a 'dispose', the team became more confident to sometimes use `subscribe()` *directly".* Future research might address this by designing experiments specifically using the users own code.

7.2 Implications

The developers using Rx in practice now have an alternative to `printf`-debugging. We recommend developers to try RxFiddle on their codebase to better understand the reactive behavior of their applicatoin, and potentially detect and verify (performance) bugs they are not aware of. At least one example of this has already occurred in practice: one of our interview subjects reported a bug[15] in the `groupBy` implementation of RxJS, which resulted in retention of subscriptions, increased memory usage and finally led to an out-of-memory exception; the subject detected the bug in practice and required extensive amount of debugging involving the Node.js debugger to trace down, but could be validated quickly when examining the life-cycle events in RxFiddle.

Contributors of RP libraries should use tools like the RxFiddle visualizer in documentation to provide executable samples, which would allow for a better learning experience, and at the same time introduces novice developers to other ways of debugging than `printf`-debugging.

7.3 Limitations

Multiple inputs and outputs. If we compare our debugger visualization to the visualization of learning tools, like RxMarbles [33] or RxViz [37], the main difference is that those tools show all input and output Observables of a single operator concurrently, for comprehension of that one operator, while RxFiddle shows one input and output Observable per Marble Diagram, part of a single full flow (a path through the graph). The choice to show a full flow allows developers to trace events from the start until the end of the flow, but restricts us in showing only a single ancestor flow per node at each vertical position, as adding a third dimension would clutter the (currently 2D) visualization. For future research it would be interesting to compare (1) the different ways Observable streams can be combined in Marble Diagrams and (2) which visualization elements can be added to explicitly show *causality* and *lineage* for events and show *durations* for subscriptions.

Edge visualization. In our graph visualization, the edges represent the dependencies and the path of the events. Nodes with multiple incoming edges *merge* the events, however users could falsely think that all event data ends up in the outgoing path: besides for *merging data* Rx also uses Observables *for timing*, as durations (`window`), as stop conditions (`takeUntil`) and as toggles (`pausable`). Different visual representations for joining paths could be explored to distinguish between using Observables for data or for timing.

Graph scalability. Debugging large reactive systems over longer periods of time can result in significantly larger Observable graphs and Marble Diagrams than is currently evaluated. During tests of RxFiddle with larger applications like RxFiddle itself and an existing Angular application the graph became too large to render in a reasonable amount of time. Besides rendering performance, a potentially even bigger issue is with communicating large graphs to the developer. We propose several extensions to RxFiddle to remedy this issue: (1) pruning the graph of old flows to show only the active flows, (2) bundling flows that have the same structure and only rendering a single instance offering a picker into the flow of interest, (3) collapsing certain parts of the graph that are local to one source file or function, (4) adding search functionality to quickly identify flows by operator or data values, (5) support navigation between code & graph.

Marble Diagram scalability. Furthermore we think that while Marble Diagrams are useful for small to medium amount of events (< 20), both better performance and better functionality would be achieved by providing a different interface for high volume flows. Above a certain threshold of events this high volume interface could be the default, offering features like (1) filtering, (2) watch expressions (to look deeper into the event's value), and advanced features like (3) histograms & (4) Fast Fourier Transform (FFT) views. Manually examining these distinct events could take a long time. In contrast, a debugger could leverage the run-time information about the events that actually occur, to provide a UI. Advanced features like histograms could help the filtering process, while FFT could offer new ways to optimize the application by doing smarter windowing, buffering and sampling later on in the chain.

Breakpoints. Placing traditional breakpoints in a reactive program stops the system from being reactive, and therefore can change the behavior of the system. This was our reason not to include breakpoints in RxFiddle. However, the behavior of breakpoints is twofold: they allow us to modify the application state by interacting with the variables in scope, but they also provide a way to be notified of an event occurrence. While the first is arguably not desirable for reactive systems, the notification property might be a good addition to RxFiddle. BIGDEBUG [22], a debugging solution for systems like Spark [48], introduces *simulated breakpoints* for this purpose. When a simulated breakpoint is reached, the execution resumes immediately and the required lineage information of the breakpoint is collected in a new independent process. Implementing this for RxFiddle is a matter of creating the right UI as the required lineage data is already available.

8. THREATS TO VALIDITY

External validity. For the interviews we selected 5 professional developers that were both available and worked on projects involving RxJS. The online experiment was open to anyone who wanted to participate, and shared publicly. These recruitment channels pose a threat to generalizability: different practices might exist in different companies, different developer communities and for different RP implementations & languages. Future work is needed on validating the debugger in these different contexts.

[15] `https://github.com/ReactiveX/rxjs/issues/2661`

Our code samples for the tasks are based on documentation samples and common use cases for Rx; RxFiddle might perform different on actual samples from practice, especially when the developer is familiar with the project or domain. The experiment consists of 2 small and 2 medium tasks; for larger tasks the effect of using the debugger could be bigger and therefore be better measurable. Still, we chose for these smaller tasks: in the limited time of the subjects they could answer only so many questions. With the limited amount of time available, we still show that a significant speed-up can be achieved in some cases. We leave it for future work to extend the experiment to include user code and larger systems.

Construct validity. We measure the time between the moment a question is displayed and the moment its correct answer is submitted. Even though our questions and code samples are short and were designed to be read quickly, still some variation is introduced by different reading speeds of subjects. A setup where the question and code can be read before the time is started can remedy this threat; but introduces the problem of *planning* when given unlimited time [28]: subjects can start planning their solution before the time starts. Furthermore, subjects might have different strategies to validate their (potentially correct) assumptions before submitting, ranging from going over the answer once more, to immediately testing the answer by submitting it. However, explicitly stating that invalid answers do not lead to penalty might introduce more guessing behavior. Future studies could use longer tasks, with preparation time to read the sample software at hand, with a wizard-like experiment interface presenting one short question at a time.

Internal validity. As a result of the recruitment method of the experiment, a mixed group of developers took part, attracting even those without Rx experience. To reduce the variation in experience that this introduces, we separately examined the results of more experienced developers.

At the time of the experiment RxFiddle was already available online for use, and furthermore some of the experiment subjects had already used RxFiddle during piloting. We mitigate this issue partially by providing a instruction video at the start of the experiment, however subjects with extensive experience with RxFiddle might bias the results.

The *subject-expectancy effect* [28] poses a validity concern, since subjects who expect a certain outcome, may behave in a way that ensures it. Our subjects had the opportunity to learn the context of the experiment and thus could be more motivated to use RxFiddle than using the traditional debugger. Our online experiment captures motivation to some extend as drop-out (defined as quiting, before having started all tasks) happens; the approximately equal drop-out in both groups (RxFiddle 56.3%, Console 63.4%), suggests no significant motivational differences. Future studies could offer subjects external motivation (e.g. by ranking contenders and gamification [13] of the experiment, or organizing a raffle among top contenders), to limit the threats introduced by motivation.

9. RELATED WORK

RP Debugging. REScala [43] is a RP library for Scala, based on Scala.React. Recently a debugger model was created for REScala, called "RP Debugging" [44], featuring a dependency graph visualization, breakpoints, a query language and performance monitoring. The debugger fully integrates with the Eclipse IDE and the Scala debugger facilities, creating a (Scala) developer experience and a feature RxFiddle currently can not offer: reactive breakpoints. However breakpoints are arguably not as useful as "simulated breakpoints" (Section7.3, Breakpoints). Furthermore, our debugger design supports multiple languages, and works outside of the IDE, in the browser environment and/or connecting to a production system. Rx has different reactive semantics and a more powerful, but also more extensive API, which includes operators acting in the time domain (`delay`, etc.). Therefore, we argue that seeing the many values in a flow over time is very valuable; RP Debugging shows the latest values at the selected time.

RP Visualization. RxMarbles [33] visualizes single Rx operators, for the purpose of learning and comprehension. Users can modify the diagrams by dragging the events and instantly see the changes reflected in the output. By using specific precoded inputs and timings the essence of the operator is made clear. In RxViz [37], Moroshko takes a similar approach, but uses code instead of prepared inputs. Where RxMarbles is limited to non higher order flows, RxViz subscribes to all inner Observables, when it detects a higher order Observable, showing them concurrently. In contrast to our work, these tools focus only on teaching the behavior of single operators.

Omniscient Debugging. Omniscient debuggers [39] are "all-knowing debuggers", that trace, store and query all events in a program execution. When storing vasts amount of program execution information, performance and efficiency becomes very much a problem and research in omniscient debuggers focus on this specifically. We also trace events of the entire execution, however in contrast to omniscient debuggers we only store trace events regarding RP data flows. The RP semantics allow us to create future optimizations, for example retaining only the active flow structure, while the flow's data is kept in a rolling buffer.

Dynamic Analysis.. The study of program execution is called "dynamic analysis" [9]. In most cases dynamic analysis involves a *post mortem* analysis, where first the program is run, collecting an execution trace, and then the trace data is analyzed to create a visualization. The various derived visualizations, like class and instance interaction graphs, function invocation histories [29], invocation views and sequence diagrams [8] show the possibility to use trace information for debugging. Arguably an on-line analysis is more useful for debugging than the standard post mortem analysis. Reiss, in reference [41], mentions the compromises that have to be made to make an on-line analysis: reduced tracing is required to not slow down the system (known as

the observer-effect) and fast analysis is required to lower the cost of getting to the visualization, to not discourage the users. In our design, we handle the same compromises as they are relevant for RP debugging too, and our JavaScript trace implementation bears resemblance to that of Program Visualiser [29].

Understanding Debugging. Debugging for general purpose languages revolves around attaching a debugger, stepping through the code, attaching code or data breakpoints, navigating along different calls in the call stack and examining variables and results of expressions [47]. However, existing research, measuring how these different tasks are part of the developers work day, found that while developers spend much time on comprehending code, they do not spend much time inside the IDE's debugger [36]. Beller et al. [2] found that only 23% of their subjects actively use the IDE's debugger, with the most common action being adding breakpoints, followed by stepping through code. The automated tooling of these studies did not measure different kinds of debugging other than using the IDE provided tools, however Beller's survey indicates that 71% also uses `printf` statements for debugging. No indication was given of any RP language and libraries used by the subjects in the study, but the observation that `printf`-debugging is common, matches our experience with debugging reactive programs.

Debugging for Program Comprehension. Both debugging and comprehension are processes in the work of programmers. Initially, comprehension was seen as a distinct step programmers had to make prior to being able to debug programs [26], but this distinction is criticized by Gilmore, saying we must view "debugging as a design activity" [21], part of creating and comprehending programs. Maalej et al. [31] interviewed professional developers and found that developers require runtime information to understand a program, and that debugging is frequently used to gather this runtime information. This supports our view that debugging is not only used for fault localization, but also for comprehension.

10. CONCLUSIONS

Observing the current debugging practices, this work shows the prevalent method for RP debugging is `printf`-debugging. To provide a better alternative, we have created a RP debugger design and presented the RxFiddle implementation, a RP debugger for RxJS, which allows developers to: (1) gain a high-level overview of the reactive data flow structure and dependencies, (2) investigate the values and life-cycle of a specific data flow, at run-time.

Through an experiment we show that RxFiddle is an viable alternative for traditional debugging and in some cases outperforms traditional debugging in terms of time spent. There are several promising directions for improving our design. Specifically scalability could be improved and different edge visualizations could be explored, to improve the usability of the tool. Furthermore, by leveraging already captured meta data about timing of events, even more insight could be provided. At the implementation level, we plan to extend RxFiddle to other members of the Rx-family of languages.

In this paper, we make the following concrete contributions:

(1) Design of a RP debugger

(2) The implementation of the debugger for RxJS, and the service RxFiddle.net, a platform for the debugger in an online environment with code sharing functionality.

In the month after the release of RxFiddle.net the site was visited by 784 people from 57 different countries. The debugger was already used by 53 developers, excluding the use inside of the experiment. During that same period 42846 interactions with the visualizations of the debugger have been recorded, such as selecting Observables or inspecting values by hovering the mouse over the event.

The debugger and the platform are open source and are available online at [1].

References

[1] H. J. Banken, *RxFiddle, release 1.0.4*, `http://github.com/hermanbanken/RxFiddle`, Jul. 2017. DOI: `10.5281/zenodo.814981`.

[2] M. Beller, N. Spruit, and A. Zaidman, "On the dichotomy of debugging behavior among programmers", 2017.

[3] G. Berry, *Real time programming: Special purpose or general purpose languages*, INRIA, 1989.

[4] S. Blackheath and A. Jones, *Functional Reactive Programming*. Manning, Jul. 2016, ISBN: 9781633430105.

[5] J. Bonér, D. Farley, R. Kuhn, and M. Thompson, *The reactive manifesto*, 2014. [Online]. Available: `http://www.reactivemanifesto.org/pdf/the-reactive-manifesto-2.0.pdf`.

[6] J. Bonér, V. Klang, and R. Kuhn, *Akka library*. [Online]. Available: `http://akka.io` (visited on 2017-06-20).

[7] L. Campbell, *Introduction to Rx*. 2012. [Online]. Available: `http://www.introtorx.com`.

[8] B. Cornelissen, A. Zaidman, D. Holten, L. Moonen, A. van Deursen, and J. J. van Wijk, "Execution trace analysis through massive sequence and circular bundle views", *Journal of Systems and Software*, vol. 81, no. 12, pp. 2252–2268, 2008.

[9] B. Cornelissen, A. Zaidman, A. Van Deursen, L. Moonen, and R. Koschke, "A systematic survey of program comprehension through dynamic analysis", *IEEE Transactions on Software Engineering*, vol. 35, no. 5, pp. 684–702, 2009.

[10] J. W. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2013.

[11] E. Czaplicki, "Elm: Concurrent frp for functional guis", *Senior thesis, Harvard University*, 2012.

[12] E. Czaplicki and S. Chong, "Asynchronous functional reactive programming for guis", in *ACM SIGPLAN Notices*, ACM, vol. 48, 2013, pp. 411–422.

[13] D. Dicheva, C. Dichev, G. Agre, and G. Angelova, "Gamification in education: A systematic mapping study.", *Educational Technology & Society*, vol. 18, no. 3, pp. 75–88, 2015.

[14] J. Edwards, "Coherent reaction", in *Proceedings of the 24th ACM SIGPLAN conference companion on Object oriented programming systems languages and applications*, ACM, 2009, pp. 925–932.

[15] C. M. Elliott, "Push-pull functional reactive programming", in *Proceedings of the 2nd ACM SIGPLAN symposium on Haskell*, ACM, 2009, pp. 25–36.

[16] C. Elliott and P. Hudak, "Functional reactive animation", in *ACM SIGPLAN Notices*, ACM, vol. 32, 1997, pp. 263–273.

[17] A. Esposito and M. Ciceri, *Reactive Programming for. NET Developers*. Packt Publishing Ltd, 2016.

[18] J. Feigenspan, C. Kästner, J. Liebig, S. Apel, and S. Hanenberg, "Measuring programming experience", in *Program Comprehension (ICPC), 2012 IEEE 20th International Conference on*, IEEE, 2012, pp. 73–82.

[19] K. Gallaba, A. Mesbah, and I. Beschastnikh, "Don't call us, we'll call you: Characterizing callbacks in javascript", in *Empirical Software Engineering and Measurement (ESEM), 2015 ACM/IEEE International Symposium on*, IEEE, 2015, pp. 1–10.

[20] E. R. Gansner, E. Koutsofios, S. C. North, and K.-P. Vo, "A technique for drawing directed graphs", *IEEE Transactions on Software Engineering*, vol. 19, no. 3, pp. 214–230, 1993.

[21] D. J. Gilmore, "Models of debugging", *Acta psychologica*, vol. 78, no. 1, pp. 151–172, 1991.

[22] M. A. Gulzar, M. Interlandi, S. Yoo, S. D. Tetali, T. Condie, T. Millstein, and M. Kim, "Bigdebug: Debugging primitives for interactive big data processing in spark", in *Proceedings of the 38th International Conference on Software Engineering*, ser. ICSE '16, Austin, Texas: ACM, 2016, pp. 784–795, ISBN: 978-1-4503-3900-1. DOI: 10.1145/2884781.2884813. [Online]. Available: http://doi.acm.org/10.1145/2884781.2884813.

[23] F. Gutierrez, "Reactive messaging", in *Spring Boot Messaging: Messaging APIs for Enterprise and Integration Solutions*. Berkeley, CA: Apress, 2017, pp. 163–178, ISBN: 978-1-4842-1224-0. DOI: 10.1007/978-1-4842-1224-0_10. [Online]. Available: http://dx.doi.org/10.1007/978-1-4842-1224-0_10.

[24] W. E. Hanson, J. W. Creswell, V. L. P. Clark, K. S. Petska, and J. D. Creswell, "Mixed methods research designs in counseling psychology.", *Journal of counseling psychology*, vol. 52, no. 2, p. 224, 2005.

[25] R. Johnson, E. Gamma, R. Helm, and J. Vlissides, "Design patterns: Elements of reusable object-oriented software", *Boston, Massachusetts: Addison-Wesley*, 1995.

[26] I. R. Katz and J. R. Anderson, "Debugging: An analysis of bug-location strategies", *Human-Computer Interaction*, vol. 3, no. 4, pp. 351–399, 1987.

[27] S. Kleinschmager and S. Hanenberg, "How to rate programming skills in programming experiments?: A preliminary, exploratory, study based on university marks, pretests, and self-estimation", in *Proceedings of the 3rd ACM SIGPLAN workshop on Evaluation and usability of programming languages and tools*, ACM, 2011, pp. 15–24.

[28] A. J. Ko, T. D. Latoza, and M. M. Burnett, "A practical guide to controlled experiments of software engineering tools with human participants", *Empirical Software Engineering*, vol. 20, no. 1, pp. 110–141, 2015.

[29] D. B. Lange, Y. Nakamura, *et al.*, "Program explorer: A program visualizer for c++.", in *Proceedings of the USENIX Conference on Object-Oriented Technologies on USENIX Conference on Object-Oriented Technologies (COOTS)*, USENIX Association, 1995, pp. 4–4.

[30] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu, "Storyflow: Tracking the evolution of stories", *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2436–2445, 2013.

[31] W. Maalej, R. Tiarks, T. Roehm, and R. Koschke, "On the comprehension of program comprehension", *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 23, no. 4, p. 31, 2014.

[32] I. Maier, T. Rompf, and M. Odersky, "Deprecating the observer pattern", Tech. Rep., 2010.

[33] A. Medeiros, *RxMarbles*, 2014. [Online]. Available: http://rxmarbles.com (visited on 2016-09-10).

[34] E. Meijer, "Subject/observer is dual to iterator", in *FIT: Fun Ideas and Thoughts at the Conference on Programming Language Design and Implementation*, 2010.

[35] L. A. Meyerovich, A. Guha, J. Baskin, G. H. Cooper, M. Greenberg, A. Bromfield, and S. Krishnamurthi, "Flapjax: A programming language for ajax applications", in *ACM SIGPLAN Notices*, ACM, vol. 44, 2009, pp. 1–20.

[36] R. Minelli, A. Mocci, and M. Lanza, "I know what you did last summer: An investigation of how developers spend their time", in *Proceedings of the 2015 IEEE 23rd International Conference on Program Comprehension*, IEEE Press, 2015, pp. 25–35.

[37] M. Moroshko, *RxViz: Animated playground for Rx Observables*, 2017. [Online]. Available: https://github.com/moroshko/rxviz (visited on 2017-06-20).

[38] T. Nurkiewicz and B. Christensen, *Reactive Programming with RxJava, Creating Asynchronous, Event-Based Applications*. O'Reilly Media, 2016.

[39] G. Pothier and É. Tanter, "Back to the future: Omniscient debugging", *IEEE software*, vol. 26, no. 6, 2009.

[40] *ReactiveX.io*. [Online]. Available: http://reactivex.io/ (visited on 2016-09-10).

[41] S. P. Reiss, "Visualizing program execution using user abstractions", in *Proceedings of the 2006 ACM symposium on Software visualization*, ACM, 2006, pp. 125–134.

[42] G. Salvaneschi, S. Amann, S. Proksch, and M. Mezini, "An empirical study on program comprehension with reactive programming", in *Proceedings of the 22Nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*, ACM, 2014, pp. 564–575.

[43] G. Salvaneschi, G. Hintz, and M. Mezini, "Rescala: Bridging between object-oriented and functional style in reactive applications", in *Proceedings of the 13th international conference on Modularity*, ACM, 2014, pp. 25–36.

[44] G. Salvaneschi and M. Mezini, "Debugging for reactive programming", in *Proceedings of the 38th International Conference on Software Engineering*, ACM, 2016, pp. 796–807.

[45] J. Siegmund, C. Kästner, J. Liebig, S. Apel, and S. Hanenberg, "Measuring and modeling programming experience", *Empirical Software Engineering*, vol. 19, no. 5, pp. 1299–1334, 2014.

[46] J. Singer, T. Lethbridge, N. Vinson, and N. Anquetil, "An examination of software engineering work practices", in *CASCON First Decade High Impact Papers*, IBM Corp., 2010, pp. 174–188.

[47] D. Spinellis, *Effective Debugging: 66 Specific Ways to Debug Software and Systems*, 1st. Addison-Wesley Professional, 2016, ISBN: 9780134394794.

[48] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker, and I. Stoica, "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing", in *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, USENIX Association, 2012, pp. 2–2.

APPENDIX

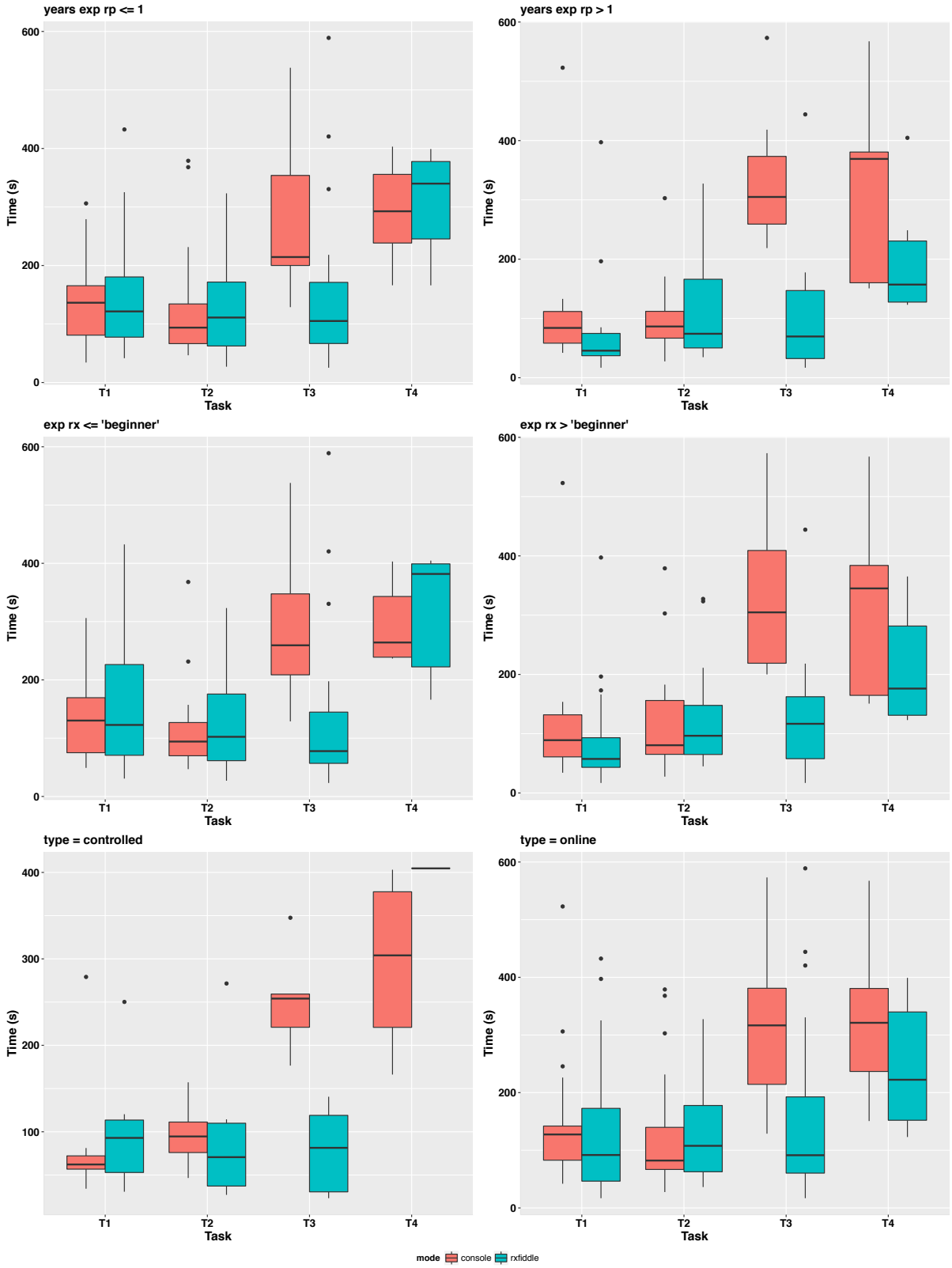


Figure 10: Time until correct answer per task, split in various groups