EXPLORATORY RESEARCH

Final Project Results Report

Deliverable ID:	D1.2
Dissemination Level:	PU
Project Acronym:	BEACON
Grant:	893100
Call:	H2020-SESAR-2019-2
Topic:	SESAR-ER4-08-2019
Consortium Coordinator:	University of Westminster
Edition date:	10 February 2023
Edition:	01.00.01
Template Edition:	02.00.05

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Authoring & Approval

Authors of the document

Name / Beneficiary	Position / Title	Date
Gérald Gurtner / UoW	Project Coordinator	14/11/2022
Graham Tanner / UoW	Project Team	11/11/2022
Tatjana Bolic / UoW	Project Team	14/11/2022

Reviewers internal to the project

Name / Beneficiary	Position / Title	Date
Graham Tanner / UoW	Project Team	11/11/2022
Tatjana Bolic / UoW	Project Team	14/11/2022
Andrew Cook / UoW	Project Team	11/11/2022

Reviewers external to the project

Name / Beneficiary	Position / Title	Date
N/A		

Approved for submission to the S3JU By - Representatives of all beneficiaries involved in the project

Name / Beneficiary	Position / Title	Date
Gérald Gurtner / UoW	Project Coordinator	14/11/2022

Rejected By - Representatives of beneficiaries involved in the project

Name and/or Beneficiary	Position / Title	Date
N/A		

Document History

Edition	Date	Status	Name / Beneficiary	Justification
01.00.00	14/11/2022	Submitted	Gérald Gurtner / UoW	First submission
01.00.01	10/02/2023	Submitted	Gérald Gurtner / UoW	Second submission after SJU comments

EUROPEAN PARTNERSHIP

Co-funded by the European Union

BEACON

BEHAVIOURAL ECONOMICS FOR ATM CONCEPTS

This deliverable is part of a project that has received funding from the SESAR3 Joint Undertaking under grant agreement No 893100 under European Union's Horizon 2020 research and innovation programme.

Abstract

The BEACON project applied behavioural economics to the user-driven prioritisation process concept across the European airspace and airport network. This final deliverable summarises the objectives, work performed and key results of the project. It shows how the project successfully designed new mechanisms for ATFM hotspot resolutions and tested them with the help of different models, including concepts from behavioural economics, in terms of economic efficiency for airlines and equity among them.

Table of Contents

List of Tables

List of Figures

1 Executive Summary

The project built on the concept of User-Driven Prioritisation Process (UDPP) developed by Eurocontrol. UDPP is a simple way for airlines to avoid impact of massive delays on their fleet, by reordering their own flights when an Air Traffic Flow Management regulation hits them. Each delayed flight has a cost for the airline, but this cost is highly contextual, depending on the type of flight, number of passengers, even time of the day. Airlines are nowadays able to protect their most important flights (i.e. the most costly) by swapping their slots with slots assigned to less important flights.

BEACON studied the feasibility of extending UDPP to allow more efficient prioritisation processes, in particular allowing slots to be exchanged between airlines. For example, low-volume airlines (those having a small number of flights in a regulation) cannot exchange slots among themselves, but may hold slots that are of real value to other companies. Exchanging these slots, against a potential compensation, should thus allow for a more efficient system from the total cost point of view.

Different types of mechanisms and credit systems were tested in BEACON. On top of their overall efficiency, a special attention was paid to issues of fairness and equity, i.e. how different airlines are impacted. The mechanisms were tested using two models (a simplified and a more extensive one) that can capture network effects in the airline cost structure, an important issue when it comes to their cost-minimisation decisions.

To properly capture the actors' behaviours, BEACON made use of behavioural economics. Non-rational behaviours like loss-aversion and hyperbolic discounting were embedded in the models in order to take them into account in the design of the new prioritisation mechanisms right from the start. Indeed, decisions featuring deviations from rationality (a common but overlooked effect in economy) may change the way the mechanisms behave with respect to a perfect rational baseline, and thus must be included from the design stage. Behavioural experiments were carried out in order to calibrate these effects, in particular using human-in-the-loop simulations.

BEACON showed that behavioural and gaming aspects may have a very important role in this context, severely decreasing the economic efficiency of inter-airlines mechanisms. The project proved that approximations and possible errors in the calculation of flight cost functions are crucial, and their effects are underestimated in the current research. Moreover, the project showed that designing an efficient inter-airline mechanism that can do better than an intra-airline one is difficult, due to the combined effects of gaming and behavioural effects, issues of approximations and errors, and relatively high level of cost efficiency of intra-airlines optimisations, which are less prone to these various issues. Finally, the project showed that equity issues are far from trivial, and that a consensus is needed on what is the actual goal of UDPP mechanisms in terms of economic efficiency and/or equity.

BEACON increased the understanding on what Behavioural Economics can add to ATM concepts elaboration and validation methodologies and deepened and broadened the concepts of prioritisation in ATM beyond UDPP and their potential impacts on Network performance. It formed a guideline and numerous recommendations for future research in this field, both from a theoretical and a practical point of view.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

2 Project Overview

2.1 Operational/Technical Context

During their operations, airspace users (AUs) are often in need of flexibility. Because of unforeseen events, including delays generated in the air traffic management (ATM) system, they sometimes need to reconfigure their operations, cancelling a flight, swapping aircraft, reordering several flights, etc., in order to reduce the impact of these events on their operational costs. On the other hand, the network needs reliability in order for air navigation service providers (ANSPs) and airports to deliver the required capacity in an efficient manner. Thus, there is a need for mechanisms creating flexibility for AUs while ensuring an adequate level of certainty for the Network Manager.

The **User-Driven Prioritisation Process** (UDPP) is a SESAR concept which aims to develop such mechanisms, by providing tools whereby the AUs prioritise their flights as a function of their business needs, in coordination with the Network Manager, in order to reduce the effects of regulations on key flights. UDPP has enhanced the possibility to swap slots of regulated flights, increased **departure flexibility** at airports, and elaborated **prioritisation mechanisms** for airlines to reorder their own flights in a regulation. These methods provide effective cost efficiency improvements to airline. Despite these successes, UDPP has also some limitations, for example low-volume users (airlines with low number of flights in the regulation) might not fully benefit from these processes.

Indeed, airlines with a low number of flights in a given regulation cannot improve their costs significantly with UDPP, since by design any slot swap happens among the slots already 'owned'[1] by the airlines. Inter-airline slot swapping should thus unlock new possibility to decrease the total cost of the system, for instance letting high volume users use the low-volume users slots against some form of retribution.

BEACON has explored a way to expand prioritisation processes by going beyond the UDPP intra-airlines slot swaps. New mechanisms have been designed to this end, going from incremental improvement over UDPP to exploratory concepts like Auctions. Mechanisms were designed in order to 'unlock' the hidden benefits of inter-airline slot swapping, exploring at the same time inter-temporal and inter-regulation slot trading. Some of the BEACON mechanisms thus naturally extends harmoniously prioritisation (or preference) settings to a larger scope geographical scope than a single airport.

To evaluate these mechanisms, BEACON adopted the perspective of **behavioural economics**, with the aim of going beyond previous modelling approaches that (unrealistically) assume perfectly rational behaviour from the AUs. BEACON developed incisive models embedding complex non-rational behaviours from AUs, using highly detailed cost models. These models have allowed us to better understand the reactions of AUs to different prioritisation processes. Moreover, BEACON has developed indicators in the domains of fairness and equity to assess the differential impact on the different types of AUs.

[1] From the First Planned First Served (FPFS) allocation.

2.2 Project Scope and Objectives

BEACON's general goal was **to design new procedures for airspace users (AUs) to better allocate their resources** (aircraft, pilots, crew, and others) in case of disruptions and evaluate the proposed procedures through new methods and tools able to consider AUs' complex behaviours, such as bounded rationality. The procedures have been designed to comply with the various requirements from AUs, such as equity. To this end, a significant amount of effort was dedicated to the exploration and definition of metrics to characterise **equity** and **fairness** in an objective way.

BEACON aimed to make significant contributions at two levels: (i) at the **methodological** level, by developing new approaches for the assessment of flight prioritisation mechanisms based on behavioural economics; (ii) at an **applied** level, by formulating and assessing new flight prioritisation mechanisms.

The project's high-level objectives were to:

- Propose a set of **improved flight prioritisation mechanisms** that expand current UDPP capabilities.
- Define **new metrics** to evaluate the **fairness** and **equity** of flight prioritisation mechanisms and validate their appropriateness with AUs.
- Quantify the impact of 'non-rational' behaviours of AUs on the outcome of the proposed mechanisms, taking advantage of the methods and tools developed in the field of **behavioural economics**.
- Integrate the insights gained from behavioural economics into an **agent-based microsimulation** model of the full ECAC network able to capture network effects. The model shall be able to compute a set of key performance indicators (KPIs), including newly developed fairness and equity KPIs, allowing a comprehensive assessment of the new UDPP mechanisms.
- Run a set of **simulation experiments** to evaluate the impact of the new UDPP mechanisms on the selected KPIs, taking into account behavioural effects, in order to analyse the advantages and the risks with respect to the current UDPP capabilities.
- Derive **guidelines and methodological recommendations** on the further development, validation and deployment of the new UDPP mechanisms that pave the way to a more harmonised and efficient flight prioritisation process across Europe.

2.3 Work Performed

The work in BEACON is focused on the reduction of the cost of delay incurred by flights involved in a regulation. Within the scope of BEACON, we used the following scope and assumptions:

- When a regulation hits, all the flights arriving within the temporal window of the regulation are part of the "regulation resolution". The regulation resolution includes a potential collaborative phase between a central planner and airlines and the computation of a final allocation. The way the central planner asks the preferences to the airlines and the way these are used to compute the final allocation is called the "mechanism".

- An objective observer (like the Network Manager (NM)) has only access to the ETAs of the flights involved, and the respective airlines to which they belong.

- Subjective observers (airlines) have only access to all internal information allowing them to prioritise their flights or form preferences[2]. The highest form of information for the airline is to have a "cost function" of each their flight, i.e. a mapping between each slot and how much cost of delay this flight would incur if it were allocated to this slot. We assume all cost functions to be independent of each other in this case.[3]

- The "central planner" (notionally represented by the NM) is in charge of asking preferences to airlines and choosing the final allocation, based said preferences and taking into account other operational constraints.

- The large-scale simulations were performed on 21 airports, among the most important ones in Europe.

The work in BEACON began with the definition of the mechanisms that would allow the airlines to swap slots among them and decrease their costs beyond the 'basic UDPP'[4] mechanism, as well as the definition of the indicators that would be used to optimise and assess the mechanisms. The project then went on to try to calibrate the behavioural models that would then be needed for the fast-time simulations, with the help of a survey and human-in-the-loop simulations. For this, and for the subsequent fast-time simulation, a library with high performance was implemented to solve some of the mechanisms. Fast-time simulations were then performed, using a small-scale model in synthetic data and a large-scale model with more realistic ones. The project dedicated a lot of effort during its course to three transversal activities: the concept assessment, where the mechanisms themselves and the concepts used for modelling were assessed; the data preparation and elaboration; and the communication activities.

2.3.1 BEACON mechanisms (WP3)

After various internal discussions and consultations with stakeholders (during a workshop in particular), three mechanisms were selected for their potential benefits and presented in D3.1 and D4.2.

ISTOP. The first one was named ISTOP, or Inter-airline Slot Trading Offer Provider. Based on UDPP and thus incremental in nature, it aims at providing inter-airline offers to pairs of airlines, beneficial for both. In order to do this, it asks airlines for their preferences, in the form of their true cost function or an approximated version of it, potentially built from an approximation of the cost function itself on relying on higher-level heuristics. Indeed, an important aspect of the user-driven preference system is that airlines in general have only an imperfect knowledge of their own costs, or more precisely that their operational decisions are not directly linked to a cost but the 'smooth' continuation of their operations. Hence, just like UDPP allows airlines to give priorities or set other parameters[5], ISTOP allows users to give higher, more general preferences for their flights. It is important to note, however,

---

[2] Here and in the following, we will use 'preferences' for any kind of indication of costs communicated from the airlines to the central planner. These can be in the form of prioritisation, full cost functions etc.

[3] In the context of the modelling activities, we assume that airlines know their true cost functions.

[4] Here and in the following, we refer to the 'basic UDPP', or just UDPP, as the possibility of intra-airline slot swapping, plus the Selective Flight Protection (SFP, see D3.1 for its definition).

[5] Like the "Time not before"/ "Time not after", see D3.1 for more details.

that airlines can directly input their cost functions into ISTOP in principle if they know them, a feature which is not available in UDPP at the moment to our knowledge[6].

The cost functions, either directly communicated or rebuilt from some parameters, are then renormalised by the central planner, by considering the maximum cost in each airline. The central planner then tries to find inter-airlines swaps between flights belonging to two airlines that are beneficial to both airlines. Here, by synthetic costs, we mean the renormalised costs built in the previous step. The central planner then communicates any beneficial swap to the airlines concerned, who may or may not accept it. This mechanism thus allows inter-airline swaps while 1) allowing airlines to in control of their final allocation, like UDPP, 2) communicate simply their cost and 3) keeping the exact details of their cost hidden thanks to the renormalisation procedure.

If ISTOP allows some degree of inter-airline swaps, it has also some limitations, due its incremental design. The most important one is that inter-airline swaps that are good from the system point of view (the total cost) may not be directly beneficial for some airlines (i.e. they will lose from the swaps in terms of cost of delay). In this case, one needs to counter-balance the local losses by granting some future advantage to the airlines.

CM. This is exactly how the second mechanism defined by BEACON, called the Credit Mechanism (CM) was designed. In this mechanism, we are not building upon the UDPP mechanism, but extending in some way ISTOP nonetheless. Indeed, in this mechanism the airlines are asked to provide some parameters that approximate their cost function, like in ISTOP. However, the cost functions built by central planner are not normalised anymore, and thus the costs are compared in absolute terms across airlines. The central planner then tries to find the best allocation from the **total cost** point of view. This best allocation is then considered as the final one, without further interaction with the airline.

However, in order to avoid that airlines conflate their costs to gain an advantage, and to allow small players to gain from such mechanism, it imposes that the parameters communicated to the central planner must be paid in a virtual currency, called credits. The airlines are indeed allowed 'default' parameters, i.e. default surrogate cost functions for which they do not have to spend credits. Any deviation from these values will require them to spend credits. Hence, airlines putting strain on system by declaring high costs will naturally transfer credits to airlines declaring low costs. Since credits can be used freely across regulations, this allows for instance low-volume users to cede the slot that would have been allocated to them with the FPFS rule if it's not important for them, get credits, which can then be spent in a future regulation to protect a more important flight.

Auction. Finally, BEACON defined a third mechanism, inspired by a previous Engage KTN[7] "catalyst" project. This mechanism is a primary Auction, where airlines bids for slots. More specifically, for each of their flight, airlines have to bid in a virtual currency, called credits again, for each of the slots of the regulation (open to this flight). The airlines can be placed bids freely, with in particular the possibility to bid a negative number of credits to gain some, with only some technical constraints.

---

[6] It is worth noting that some of the work performed during BEACON can be used exactly for this purpose. In particular, a library (Hotspot) has been built to allow users to form prioritisations and protection requests on their flights based on their true cost function, hence fully automatising the UDPP process if airlines know their cost functions.

[7] Engage, the SESAR 2020 Knowledge Transfer Network (https://engagektn.com/).

The bids are resolved slot by slot by the central planner, with the flight with the highest bid getting the corresponding slot. The credits are then paid to the central planner, who redistributes them equally among airlines in order to have a constant number of credits in the system. The airlines cannot modify their position in the queue after that.

This auction mechanism solves the difficult issue of having the airlines communicating some approximated costs to the central planner. Instead, this form of trading ensures that airlines pay only the slots for it is worth for them. Obviously, a major drawback is to have to choose the bids for each airline and each slot, which means that in practice some degree of automation is needed.

Different scenarios were defined in WP3 in order to test the mechanisms in different setups, including dimensions like traffic conditions and agents' behaviours. This list of scenarios, together with priorities was refined and used in WP4 and WP5 by the two models, based on on-going results. The final selection of the scenarios used in the model was driven by the scientific plan to explore the behaviour of the models and obtain comprehensive results leading to the completion of the objectives.

2.3.2 Indicators and cost of delay update (WP3)

The definition stage of the project also involved a comprehensive study on the cost of delay of airlines. Such a study already existed [10], but was dated from 2015. BEACON took the opportunity that this cost of delay model was at the heart of the modelling process to update it.

The key updates to the reference values are included, updating the costs based on market changes since 2015, newly adding important consideration of ***airport curfew costs***, and, *inter alia*, updating the assessment of the passenger cost of delay driven by Regulation 261 trends. Since this was reviewed by a range of airspace user stakeholders, with their helpful feedback and that of the SJU taken into account in the final deliverable, the University of Westminster proposes to share this deliverable with the PRU and EUROCONTROL ('Standard Inputs for Economic Analyses') team, who currently use Westminster's cost of delay reporting.

Another important feature of BEACON is to capture the fair, or equitable, dimension of the mechanisms tested during the project, or even beyond the project. Indeed, it was clear from the beginning of the project and again during the feedback gathered during its course that the acceptability of a regulation resolution mechanism does not rely on its pure economic efficiency, is also related on the differential advantages it brings to airlines. For instance, a mechanism produced a very low cost of delay but always advantaging the same airline, or the same type of airlines, will not be accepted. In order to study this aspect, the project defined and used three different indicators (some of them are defined in D3.1, but their most complete description is in D5.2) for the 'equity' in the impact of the mechanisms.

2.3.3 Survey and human-in-the-loop simulations (WP4 and WP5)

The first stage of the project also saw the organisation of a survey related to behavioural economics. Indeed, one of the stated goals of BEACON is to use behavioural economics to have a better assessment of the impact of the mechanisms. **Behavioural economics** is a rapidly emerging field, delivering important insights to the analysis of human behaviour and decision-making through several disciplines of study - psychology, neuroscience, economics and decision science. Tens of thousands of decisions are made by individuals every day. These decisions are subject to mental shortcuts (heuristics) and biases (Gilovich, Griffin and Kahneman, 2002 [9]). In the context of BEACON, we used the co-called Prospect Theory and Hyperbolic Discounting models to simulate the behaviours of airlines making decisions during the hotspot resolution.

The survey we launched was supposed to help us calibrate the behavioural models used in the rest of the project. Numerous relevant members of the industry were contacted to fill the survey, during three different campaigns of dissemination. However, the survey received very few responses, leading the consortium to rely on a literature review instead to find adequate values for the coefficients entering the two models (prospect theory and hyperbolic discounting). The survey and the results of the literature review can be found in D4.1.

Related to this line of work, we also launched human-in-the-loop simulations. These simulations were performed by developing a human-machine interface (HMI) dedicated to communicating with the Mercury simulator (used in fast-time simulations too), and let a human play the role of an airline involved in a regulation that needs to make decision on the values of the parameters to communicate to the central planner in the context of two mechanisms: ISTOP and the CM. These simulations were designed both to have a further calibration of behavioural models and get some feedback on the mechanisms, but the latter motivation proved to be much more fruitful in the end. The work invested in Mercury to adapt it to real time simulations proved very useful and some members of the consortium plan to use the simulator for such use in the future again.

2.3.4 The Hotspot library and the 'simplified' simulations (WP4)

The second stage of the project was dedicated to the modelling process leading to the estimation of indicators related to the impact of the mechanisms on the airline costs. For this, we used two separate models, a small scale one developed in WP4 and the Mercury simulator, already existing but expanded and enhanced for BEACON.

The first model was focused on producing results on a simplified, small-scale and synthetic dataset, in order to have a first idea of the magnitude of the indicators involved. Directly inspired from the model developed during an Engage KTN "catalyst" project, this model implemented the three BEACON mechanisms with simple behavioural rules for airline agents, that included prospect theory and hyperbolic discounting models. The model included five airports and nine en-route sectors, took into account capacity-demand balance, passenger connections for flights, and solved regulations at arrival airports. The behavioural models were simplified in order to allow quick computations but showed the first effects of behavioural economics on the impact of mechanisms. The presentation of the model and the results obtained are all presented in D4.2.

In order to run the model, a library named "Hotspot" was developed in WP4 too, which was also used in WP5 (see next section). This library, written partly in C and partly in Python, was written to allow to resolve efficiency some mechanisms, in particular ISTOP and the CM, plus some mechanisms used as benchmarks (including the basic UDPP). Even though it was designed with the models from WP4 and WP5 in mind, the library is stand-alone, usable for other purposes and whose source code may be opened at the end of the project. Therefore, it represents a major piece of work from BEACON and can be considered as one of the major outputs on the methodological side.

2.3.5 Large scale or 'realistic' simulations (WP5)

The second model, reusing the existing simulator Mercury, was geared towards a more realistic estimation of the impact of the mechanisms. Using historical data for regulation, flight, and passenger data, it was used to create intermediate datasets of regulations on which small 'games' were played by artificial agents. Hence, the statistical properties of the distributions of delays, flight per regulation etc. reflect actual distributions and led to a more realistic assessment of the mechanisms than in WP4.

These agents were designed to be able to have different behaviours, in order to test the mechanisms under different circumstances. Three behaviours were defined as: the 'honest' one, which the airline communicated the best approximation of its costs to the central planner; the 'rational' one, in which the airline tried more aggressively to communicate parameters to the central planner that would increase its expected profit (decrease the expected cost of delay of the final allocation); and the 'bounded' one, similar to the rational one, but trying to increase its expected utility instead, distinct from the pure profit because of behavioural biases.

In D5.1, the behaviours of agents, the cost approximation linked to the mechanism, and calibration issues were explored in detail, using regulation applied on one airport as examples. In D5.2 instead, the focus was on computing more realistic indicators, with more airports involved, and explore the behaviour of the mechanisms as a function of the size of the regulation, the size of the airlines involved etc. The emphasis was thus very much not only on the pure economic efficiency of the mechanisms, but also on their differential impact they have on different types of airlines, in other words on equity.

2.3.6 Concept assessment (WP6)

On top of the technical work above, the consortium spent some important effort to assess the concepts used in BEACON, and more generally to reflect on the way to extend UDPP to an inter-airline, inter-temporal, and inter-regulation mechanism.

We thus explored the barriers to implementation of potential mechanisms, the relationship between efficiency, usability, and automation for these mechanisms, the theoretical and practical limits to the efficiency of the mechanisms, the issues related to data sensitivity, virtual currencies etc.

All these reflections are compiled in D6.1 and D6.2 and are aimed at providing a basis for future research on the subject. There fuelled directly the conclusions of the project and the plans for future research?

2.3.7 Data preparation and elaboration (WP2)

Another important transversal activity that was carried out during the project is the data acquisition, preparation and elaboration. Indeed, especially in the context of WP5, the project needed reliable historical data on which to base the models used for the mechanism assessment.

Data related to passengers, schedules, and flight data were thus acquired. The data were cleaned, merged, and compiled in a database accessible to the consortium. The database was used extensively for descriptive analyses, as input for Mercury, and for the elaboration of intermediate datasets useful for some experiments.

2.3.8 Dissemination and exploitation (WP7)

Finally, the project was involved in several events related to dissemination and exploitation, including workshops (one of which was organised by BEACON) and conferences.

Conferences. The project has been disseminated at two editions of the SESAR Innovation Days and the AGIFORS Airline Operations Study Group conference:

- **10th SESAR Innovation Days**, 07-10 December 2020 (1060 participants) – BEACON poster presented, plus video and one minute teaser (https://www.youtube.com/watch?v=XxZ02Rg-XCo).
- **11th SESAR Innovation Days**, 07-09 December 2021 (758 participants) – BEACON poster presented, plus video and one minute teaser (https://www.youtube.com/watch?v=p2gTSOqnA9U).
- **AGIFORS Airline Operations Study Group conference**, 13-16 July 2021 (approx. 100 participants) – BEACON presentation 'Credit-based mechanisms for user-driven prioritisation during ATFM regulations' (abstract https://agifors.org/page-18323).
- **AGIFORS 62nd symposium,** September 2022 – BEACON presentation "Inter Airline Slot Trade Opportunities Provider (ISTOP)" (presentation https://publications.agifors.org/documents/1-04%20Andrea%20Gasparin%20ISTOP.pdf). Won the "best innovation" paper award.

Additionally, a paper has been submitted to the forthcoming 12th SESAR Innovation Days conference (2022).

Workshops. A stakeholder workshop was organised with the BEACON Advisory Board on 11 November 2020 in a virtual setting to discuss BEACON's mechanisms as well as any barriers to implementation. Experts from HungaroControl, NATS, Air Baltic, El Al, Hop!, Schiphol airport, EUROCONTROL (as PJ07-W2 leader) and Frequentis (as SlotMachine coordinator) participated. Members of the consortium have participated in other workshops, for example, joining as a panellist in session 2 of the Engage KTN workshop that focused on 'Economic incentives for future ATM implementation' (held virtually on 21 June 2021 with 67 participants).

A final dissemination workshop was also organised in January 2023, after the end of the project, to present the results to the wider community. The workshop had approximatively 15 participants and focused on the problem of estimating cost functions from airlines and designing new mechanisms that are better than UDPP.

Website. The BEACON website launched in October 2020 (https://www.beacon-sesar.eu/), with content summarising the project, the consortium, key activities and hosting the public deliverables.

Social media. BEACON's Twitter account launched in December 2020 (twitter.com/H2020Beacon), attracting 38 followers. Approximately 20 Tweets have been sent, with content including an introduction to the project, the poster at the 10th SESAR Innovation Days and promotion of BEACON's on-line survey. A LinkedIn account (https://www.linkedin.com/in/beacon-project/) has also been used to support project communication.

Article. The project was included in the December 2020 edition of the EUROCONTROL Innovation Hub newsletter.

Human-in-the-loop simulations (HITL). The HITL sessions interacting with Advisory Board members and dispatchers that agreed to participate in HITL (eight sessions). March/April 2022.

Survey for behavioural economics. Two major disseminations to EUFALDA and IATA, and several dissemination to the Advisory Board. April 2021-March 2022.

Further reporting on these activities can be found in D7.1.

2.4 Key Project Results

The human-in-the-loop simulations (HITL) were performed, with the goal to compare the behaviour of real dispatchers against their modelled counterparts. In particular, this task focused on estimating the importance of behaviours deviating from 'rationality'[8]. To carry out this exercise the development of a human-machine-interface (HMI) was necessary, and the adaptation of the Mercury simulator[9] to enable the interaction between the participants and the simulator. The HMI enabled the participants to play the dispatcher role which could have been otherwise played by an artificial intelligence subroutine in the simulator.

One of the findings important for the development of the BEACON mechanisms is that none of the participants had issues with exchanging slots with other airlines. Furthermore, they did not mind not having access to the results for other airlines, as they were focusing on the best solutions they could provide for their own airline.

As an illustration of our most important findings obtained with the models, we show in Figure 1 and Figure 2 the efficiency of the different mechanisms tested, with some benchmarks, and the corresponding equity indicators. The first figure shows the efficiency (relative cost savings) of the 'basic' UDPP, the ISTOP mechanism (on top of UDPP), the NNBOUND benchmark (minimum total cost with no airline losing from the allocation with respect to FPFS), GLOBAL benchmark (minimum total cost), and the CM, as a result of the simulations from D5.2. The simulations are performed with different agents types: agents providing their true costs ('exact'), agents providing their best approximated costs ('honest'), agents providing costs so that their expected gains are maximised ('rational') and agents providing cost so that their expected utility/prospect is maximised ('bounded' and 'bounded-simple').

[8] As defined in previous deliverable (and in the economic/game theory field in general), rationality is defined as the course of action that would lead to the best outcome for a given agent, given a situation and the level of information accessible to the agent.

[9] As reminder (it is explained more in details in other deliverables, see D3.1, D5.1, and D5.2), the Mercury simulator is a model developed over several years to compute advanced indicators on the air (and later ground) transportation system, such as the distribution of delays for passengers, flights, etc. It uses a strong agent-based paradigm, particularly suited to heterogeneous systems.

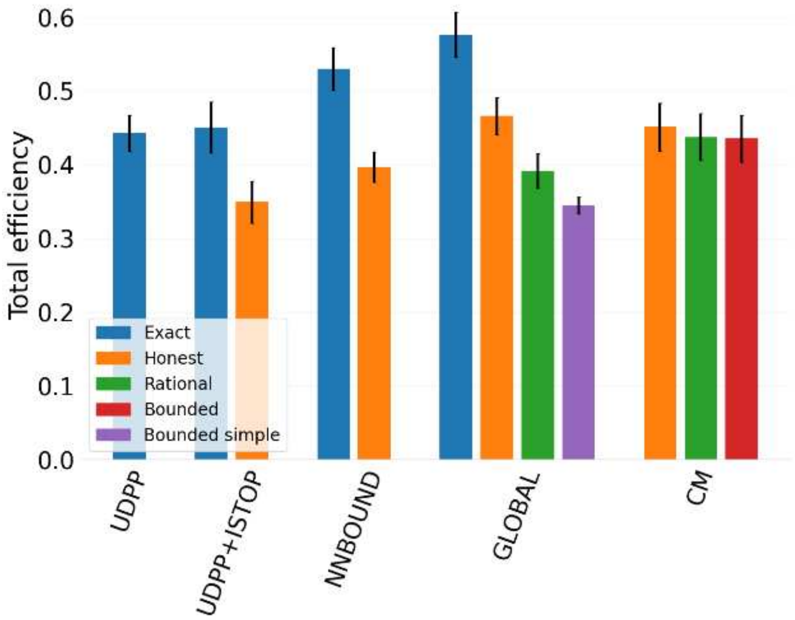

Figure 1: Efficiency of mechanisms with different agent types implemented. Error bars are standard errors.

Figure 2 shows the corresponding equity metrics for the same situations. The first plot shows EQ1, related to the average on the differences in absolute saved costs among airlines, the second one shows EQ2, related to the average on the differences in cost saved per flight among airlines, and the third one shows EQ3, related to the average on the differences in relative saved cost among airlines.

Some of the most important key findings from BEACON are directly linked to these figures and are explained thereafter.

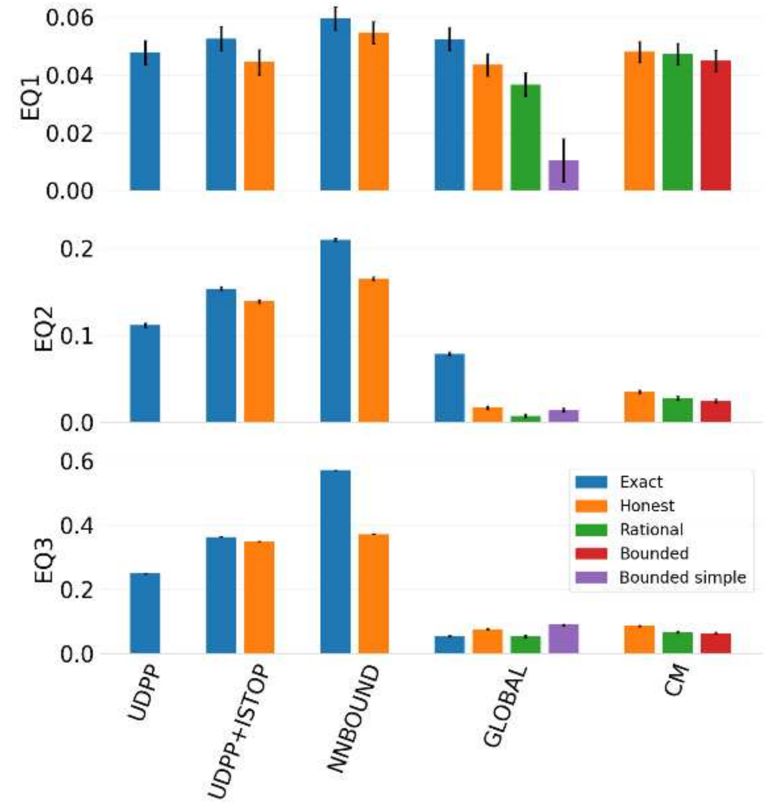

Figure 2: Equity indicators related to absolute saved cost (EQ1), saved cost per flight (EQ2), and relative saved costs (EQ3) for the different mechanisms and type of agents implemented.

The first key result is the fact that the 'basic' UDPP mechanism seems to have already a very high efficiency compared to the other mechanisms. We found that UDPP can reach around 45% of reduction of cost[10] in average with respect to the FPFS costs. This figure can be compared to the maximum theoretical one, around 55% in average in our setup, but more importantly to the realistic results of the ISTOP mechanism (around 35%) and the CM mechanism (around 45% as well). Given the additional issues related to approximations and behavioural effects (see below), this fact has to be taken very seriously: it means that the theoretical efficiency of any new mechanism that aims at solving the regulation problem needs to be very high in order to hope being better than the 'simple' UDPP mechanism.

Second, we proved that the integration of gaming and behavioural effects in the modelling process seems unavoidable to estimate properly the indicator. Indeed, the example of the GLOBAL benchmark showed that the efficiency of the mechanisms can be several impaired by the presence of these behavioural effects (with a drop of around 8 percentage points). Since we do not expect such effects in UDPP, it seems hard to beat the UDPP efficiency in a realistic setting. As a consequence, we suggest

[10] This figure assumes no error on the computation of costs on the airline side. If such errors existed, which is likely, they would impact the other mechanisms as well. It also assumes no behavioural effects, that we find extremely unlikely in this case (because the airlines do not play against each other in this case), as least ***if*** the costs are reasonably well known. If not, the airlines may rely on various heuristics that will decrease the efficiency of the mechanism. However, it is important to note that, once again, if the airlines rely on these heuristics, the other mechanisms are likely to be very impacted as well.

using behavioural economics and game theory frameworks from the design stage when building new regulation resolution mechanisms to assess the potential impact. We also proved that both prospect theory and hyperbolic discounting should be taken into account, since they seem to both have an effect on the results.

Also related to the efficiency of some mechanisms, we also proved that costs should be approximated carefully. Indeed, any deviation from true costs seems to have quite harsh adverse impact on the efficiency of mechanisms relying on them (see differences between ‘exact’ and ‘honest’ in the figures). Since approximations are unavoidable in some mechanisms, even with perfect agents, and even worse, given than approximations are probably unavoidable within the airlines to compute their costs, this issue should be taken into account early in the design process too.

However, we also proved that the CM may be a good candidate for future research on inter-airline slot swapping. Indeed, some participants to the human-the-loop simulations found it quite intuitive from the trade-off point of view (putting pressure on the system costs credits, releasing it provides some), its efficiency was similar at least to UDPP and it seems less sensitive to gaming and behavioural effects than other mechanisms. The only disappointing feature is its low score in terms of equity, even though it was designed to try to help low-cost users to make gains.

Auction has proved to be an interesting mechanism as well, with some theoretical efficiency that might be higher that the CM (it was only tested in WP4, not with the full model in WP5). However, it is also clear that using this mechanism manually is very tedious, due to the number of bids to be completed for each regulation (one per slot per flight). Hence, any further research on this mechanism should reflect on the possible degree of automation to be used with this mechanism.

Finally, we want to highlight the need for a consensus on which indicators to consider for improving the system. We have shown, with three different types of equity indicators, that some mechanisms may prove equitable according to one indicator but much less according to another (see CM and GLOBAL in Figure 2). We have also shown the limits of some indicators like the relative savings (called efficiency in BEACON) which present some serious limitations.

2.5 Project Deliverables

Table 1: Project Deliverables

Reference	Title	Delivery Date[11]	Dissemination Level[12]
Description			
D1.1 [1]	Project Management Plan	15/02/2021	Confidential
This document is the Project Management Plan (PMP) of the SESAR 2020 Exploratory Research action BEACON. The PMP documents the management plan and procedures, complementing the project information provided in the Grant Agreement Description of Action with additional detail.			

[11] Delivery data of latest edition.

[12] Public or Confidential.

Reference	Title	Delivery Date[11]	Dissemination Level[12]
Description			
D1.2 [2]	Final Project Results Report	14/11/2022	Public
The final project results report summarises the work carried out by the project during its course, the technical results obtained, how they fit in the SESAR strategic programme, as well as lessons learned and potential future development headings on the topic			
D2.1 [3]	Data Management Plan	03/02/2021	Public
This deliverable presents the approach of BEACON for the data management, and details on the data sources considered.			
D2.2 [4]	Database structure and data elaboration	03/08/2022	Public
This technical deliverable describes the database used in BEACON, including the data structure adopted, data sources, and data elaborations that were needed for the data analyses in other WPs. All data have been used to set up and run the simulations defined by tasks 4.1 and 5.2. The deliverable includes a diagram of the relational database and the full table of the data considered with a description of their usage. Furthermore, human-in-the-loop simulations were performed to collect the data for calibration of behavioural parameters to be included in the final simulation. The data collected in the human-in-the-loop is described.			
D3.1 [5]	High-level modelling requirements	26/03/2021	Public
This document provides all the necessary high-level modelling requirements needed for the proper development of the BEACON project. Firstly, it defines an assessment framework for the performance evaluation of the different flight prioritisations mechanisms selected. The suggested framework is based on a combination of desk research and consultation with different air traffic management (ATM) stakeholder representatives. Secondly, it provides a detailed and exhaustive review of the flight prioritisation and trajectory allocation mechanisms proposed in the literature, ultimately identifying and selecting a final set of promising concepts to improve the performance of the ATM system in situations of demand-capacity constraints, to be included in BEACON simulations. Finally, it describes the different variables and parameters that are part of the possible simulation scenarios and selects the potentially most interesting combinations to measure the performance of the proposed prioritisation mechanisms.			
D3.2 [6]	Industry briefing on updates to the European cost of delay	15/11/2021	Public
The cost of delay is a critical input for the assessment of flight prioritisation mechanisms and airspace user decision-making. The University of Westminster produces the standard industry reference work for European cost of delay assessment used, for example, by airspace users, ANSPs, in performance assessment by the Performance Review Unit (PRU) and cost benefit analysis in SESAR. This deliverable includes key updates that will feed these reference values, to update the costs based on market changes since 2015, newly adding important consideration of airport curfew costs, and, *inter alia*, updating the assessment of the passenger cost of delay driven by Regulation 261 trends. After further consulting with the University of Westminster's stakeholder base, these results will be made available to the wider community and shared with the PRU and SESAR.			
D4.1 [7]	Behavioural Model Parameter Calibration	14/12/2021	Public
This document provides the necessary behavioural background and calibration for the behavioural modelling requirements of the BEACON project. Firstly, it introduces and defines key behavioural concepts that are suited for modelling purposes. These are, in particular, Prospect Theory and Hyperbolic Discounting. The suggested concepts are explained in detail and their use case parameter calibrations are summarised. Secondly, it provides the methodology used to estimate the relevant parameters with the help of an on-line survey that has been distributed to a wide range of relevant participants such as flight dispatchers. While we have,			

Reference	Title	Delivery Date[11]	Dissemination Level[12]
Description			
unfortunately, not been able to collect sufficient responses to our survey, to directly estimate the parameters, we will make suggestions on the parameters to use based on existing literature and parameters calibrations.			
D4.2 [8]	Final model results	17/08/2022	Public
This deliverable presents the small-scale agent-based simulation model used in WP4 to analyse the performance of different flight prioritisation mechanisms selected in WP3. The document contains a detailed description of the model and its most important components. Additionally, it offers the final model results and conclusions obtained with the set of simulation experiments.			
D5.1 [9]	First tactical model and results	10/11/2022	Public
The deliverable presents the results obtained with the help of the Mercury simulator in order to estimate the efficiency and equitability of hotspot resolution mechanisms defined in D3.1 and tested on a small-scale simulator in D4.2. To this end, fast-time games are defined, with a central optimiser implementing the mechanisms computing the final flight/slot allocation. The game is played by agents representing the airlines present in the regulation, and who are tasked with sending information to the central optimiser regarding their own costs, in order for the latter to find the best possible allocation cost-wise. The deliverable defines the various games possible, combining different types of central optimisers and different types of agents, taking into account various degrees of rationality and behavioural biases for the decisions of the latter. The deliverable presents a theoretical framework for these behaviours, highly simplified but implementable in simple simulations.			
D5.2 [10]	Final tactical model and results	07/11/2022	Public
This deliverable presents the final results obtained with the model developed in BEACON to estimate the impact of new mechanisms for air traffic flow (ATFM) regulation resolutions using in particular behavioural economics as part of the modelling process. Compared to the previous deliverable, it presents similar analyses, obtained on an extended geographical scope and more focused on understanding the multi-dimensional impact of the mechanisms. The results demonstrate that the impact on different airlines and airports is heterogeneous. Furthermore, it explores the relationship existing between regulation features, airline characteristics, and indicators like cost saved per flight. High-level conclusions on the impact of the different mechanisms are given, in terms of absolute and relative savings as well as in terms of equity and fairness. Finally, the human-in-the-loop simulations performed during the project are described.			
D6.1 [11]	Intermediate concept assessment report	26/03/2021	Public
This technical report presents an interim synthesis of the stakeholders' input and the specifications stemming from internal discussions and the stakeholder workshop. It presents the main issues related to the concepts attached to the mechanisms selected in D3.1, for instance in terms of market mechanism design. Following the output of these discussions, some specifications for the project, and sometimes for the models, are laid out.			
D6.2 [12]	Final concept assessment report	07/11/2022	Public
In this deliverable we present the final concept assessment analysis. We review the stakeholders' feedback collected during the human-in-the-loop simulation (HITL) activity, concentrating on the benefits as well as the weaknesses of the proposed mechanisms, and with a particular focus on their feasibility in terms of real scenario implementation. We will compare the HITL considerations with the various simulations drawing final conclusions and recommendations for future research and how this could take advantage of the lessons learned during the project.			
D7.1 [13]	Final dissemination and communication report	30/11/2022	Public

Reference	Title	Delivery Date[11]	Dissemination Level[12]
Description			
The final dissemination and communication report presents the events and actions attended or performed by the team in order to make the goals, the methodology, and the results of the project known to the wider community.			
D8.1 [14]	H - Requirement No. 1	19/03/2021	Confidential
This ethics deliverable documents how the participants for undertaking surveys and participating in simulations are selected and recruited, the processes used to obtain valid and appropriate consent, which is given freely and independently, in the absence of coercion, in light of information provided to the participant.			
D8.2 [15]	POPD - Requirement No. 5	19/03/2021	Confidential
This ethics deliverable documents how the data acquired through surveys, and the simulation exercise, will be protected by the application of anonymisation (for survey exercise) and pseudo anonymisation (for simulation exercise) techniques by the consortium. For information on the target demographic of the surveys and experiments, please refer to D8.1.			
D8.3 [16]	POPD - Requirement No. 7	19/03/2021	Confidential
This ethics deliverable describes that the data the consortium intends to process is relevant and limited to the purposes of the research project (in accordance with the ‘data minimisation’ principle). Furthermore, the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants are described.			

EUROPEAN PARTNERSHIP

Co-funded by the European Union

3 Links to SESAR Programme

3.1 Contribution to the ATM Master Plan

The BEACON project formulated the following solution, termed as BEACON solution:

"The solution consists of a new approach to ATM solution design by including concepts from behavioural economics into micro-simulators aimed at assessing the impact of new mechanisms and solutions in ATM. The new approach targets high-complexity solutions that may have knock-on effects on the air transportation system (including other solutions) and require extensive modelling effort before deployment. It focuses on a very fine-grained description of the system and actors, coupled with realistic contextual behavioural rules for the actors making decision. The added value lies in the inclusion of behavioural economics matters into modelling and as such improves the performance framework, enabling more accurate and complex simulation-based performance assessments across a variety of key performance indicators."

The BEACON solution is closely linked to the two operational improvements (OIs), but *did not work directly on the further development* of these OIs. Regarding AUO-0106, BEACON used the Fleet Delay Reordering (FDR) and Selective Flight Protection (SFP) features descriptions as benchmark and to feed the design of ISTOP mechanism. The BEACON project performed in-depth analysis of the occurrences and impact of various mechanisms on the low-volume users. These results could be used for the further work on the AUO-0107 OI.

The BEACON solution aims at reaching TRL1 after the review from the SJU at the Maturity Gate meeting.

Table 2: Project Maturity

Code	Name	Project contribution	Maturity at project start	Maturity at project end
BEACON-SOL	Inclusion of behavioural economics into ATM solution design	Development of the concept of the solution	TRL0	TRL1

3.2 Maturity Assessment

Table 3: ER Fund / AO Research Maturity Assessment

ID	Criteria	Satisfaction	Rationale - Link to deliverables - Comments
TRL1.1	Has the ATM problem/challenge/need(s) that innovation would contribute to solve been identified? - Where does the problem lie? - Has the ATM problem/challenge/need(s) been quantified that justify the research done? Note: an initial estimation is sufficient	Achieved	The ATM challenge has been formulated as: "During their operations, AUs often need flexibility. Because of unforeseen events, including delays generated in the air traffic management (ATM) system, they sometimes need to reconfigure their operations, cancelling a flight, swapping aircraft, reordering several flights, etc., in order to reduce the impact of these events on their operational costs. On the other hand, the network needs reliability in order for ANSPs and airports to deliver the required capacity in an efficient manner. Thus, there is a need for mechanisms creating flexibility for AUs while ensuring an adequate level of certainty for the Network Manager. The UDPP, a SESAR concept, aims to develop such mechanisms, by providing tools whereby the AUs prioritise their flights as a function of their business needs, in coordination with the Network Manager, in order to reduce the effects of regulations on key flights. Despite various successes, UDPP has some limitations, for example low-volume users (airlines with low number of flights in the regulation) might not fully benefit from these processes. BEACON's general goal was to design new procedures for AUs to better allocate their resources in case of disruptions and evaluate the proposed procedures through new methods and tools able to consider AUs' *complex behaviours*, such as bounded rationality. The procedures have been designed to comply with the various requirements from AUs, such as equity. To this end, a significant amount of effort was dedicated to the exploration and definition of metrics to characterise equity and fairness in an objective way." The ATM challenge is detailed in deliverable D3.1. Th BEACON mechanisms are described in deliverables D3.1 and D4.2. The incidence of low-volume users in regulations across different European airports is reported in D5.2.

TRL1.2	Have the solutions (concepts/capabilities/methodologies) under research been defined and described?	Achieved	The BEACON solution has been defined as "The solution consists of a new approach to ATM solution design by including concepts from behavioural economics into micro-simulators aimed at assessing the impact of new mechanisms and solutions in ATM. The new approach targets high-complexity solutions that may have knock-on effects on the air transportation system (including other solutions) and require extensive modelling effort before deployment. It focuses on a very fine-grained description of the system and actors, coupled with realistic contextual behavioural rules for the actors making decision. The added value lies in the inclusion of behavioural economics matters into modelling and as such improves the performance framework, enabling more accurate and complex simulation-based performance assessments across a variety of key performance indicators."
TRL1.3	Have assumptions applicable for the innovative concept/technology been documented?	Achieved	The initial assumptions have been documented in deliverables D3.1 and in the experimental plan. The final assumptions regard fine-tuning of various aspects of flight prioritisation mechanisms, like cost approximation and inclusion of behavioural economics aspects in the development of the mechanisms. As mentioned in section 2.4, and described in detail in deliverable D5.2, costs should be approximated carefully, as any deviation from true costs seems to have quite adverse impact on the efficiency of mechanisms relying on them. Further, the integration of gaming and behavioural effects in the modelling process is necessary to estimate properly the efficiency and equity indicators. Thus, we suggest using behavioural economics and game theory frameworks from the design stage when building new regulation resolution mechanisms to assess the potential impact.

TRL1.4	Have the research hypothesis been formulated and documented?	Achieved	BEACON started with the hypothesis that inter-airline flight swapping represented an untapped potential to decrease costs related to ATFM delays. If this hypothesis were true, it would have meant: • That one can design an inter-airline mechanism better than UDPP, • That this mechanism should be tested in fast-time simulations and show a significant improvement over UDPP, • That the mechanism should be tested taking into account effects inherent to inter-airline mechanisms but which are not present in UDPP. For this reason, we formulated the hypothesis that in inter-airline mechanisms, behavioural mechanisms would play a role and thus that it was necessary to model them inside the simulations. Moreover, UDPP was constrained in efficiency by the number of flights available for slot swapping inside regulations, which meant that it was not fair for low-volume users. Hence, we formulated the hypothesis that an inter-airline mechanism featuring a compensation scheme for airline leaving “good slots” would allow to increase efficiency AND fairness at the same time. These are hypotheses are described and discussed more in detail in deliverables D1.1 and the experimental plan (attached to this document).
TRL1.5	Do the obtained results from the fundamental research activities suggest innovative solutions (e.g. concepts/methodologies/capabilities? - What are these new concepts/methodologies/capabilities? - Can they be technically implemented?	Achieved	BEACON results suggest innovative methodologies to implement and use in the design of prioritisation mechanisms, as well as the design of new mechanisms themselves. The methodologies revolve around the cost approximation and inclusion of behavioural economics/gaming methods at the initial design stages of any new mechanism that includes intervention by human decision makers (and cost assessments). All the mechanisms require the cost of delay for the flights involved in the regulations. Thus, the cost modelling and cost approximation are of paramount importance. Deliverable D3.2 updated the values for cost of delay of the de facto industry standard document “European airline delay cost reference values” to the 2019 values. The usage of the values is described in detail in the original document [10].

			Furthermore, in order to facilitate the simulation modelling, a Hotspot library was created. The library is stand-alone, usable for other purposes, the source code of which will be opened after the end of the project. Therefore, it represents a major piece of work from BEACON and can be considered as one of the major outputs on the methodological side. To carry out this exercise the development of a human-machine-interface (HMI) was developed, as well as the adaptation of the Mercury simulator to enable the interaction between the participants and the simulator. The HMI enabled the participants to play the dispatcher role which could have been otherwise played by an artificial intelligence subroutine in the simulator. This enabled successful HITL runs. The methods for implementation of the gaming strategies and types of agent behaviour are described in deliverables D4.2, D5.1, and D5.2. A discussion on the technical and procedural issues that an implementation of the new mechanism would face can be found in D6.1 and D6.2
TRL1.6	Have the potential strengths and benefits of the solution identified and assessed? - Qualitative assessment on potential benefits. This will help orientate future validation activities. Optional: It may be that quantitative information already exists, in which case it should be used.	Achieved	The 'basic' UDPP tested here provides high efficiency with respect to the maximum possible gains, which means that any new mechanisms for hotspot resolutions need to be carefully designed on order to obtain additional gains. Inter-airline slot swapping seems to be accepted by the stakeholders. During the HITL, none of the participants had issues with exchanging slots with other airlines, and they did not mind not having access to the results for other airlines, as they were focusing on the best solutions they could provide for their own airline. We have shown that agents seeking their own profit over anything else (a.k.a rationality) can severely impair any mechanism. Furthermore, behavioural effects may distort the players' decisions to such an extent that the gains made on top of UDPP can be lost. As a consequence, we recommend having any new mechanisms tested at the design stage taking into account both gaming and behavioural effects. Finally, we have shown, with three different types of equity indicators, that some mechanisms may prove equitable according to one indicator but much less according to another.

			The potential strengths and benefits of the solution are described in section 2.4, and more details are given in deliverables D6.1 and D6.2.
TRL1.7	Have the potential limitations, weaknesses and constraints of the solution under research been identified and assessed? - The solution under research may be bound by certain constraints, such as time, geographical location, environment, cost of solutions or others. - Qualitative assessment on potential limitations. This will help orientate future validation activities. Optional: It may be that quantitative information already exists, in which case it may be used.	Achieved	By implementing rational and bounded agents, the consortium found some serious theoretical issues, that cannot be overcome only by using more powerful machines. Reinforcement learning is thought to be a good framework to have a better estimation, even though the solvability of the model would still be an open question (but this time related to the computational power). Due to the complexity of solving a multi-partite mechanism where airlines compete for the best slots, it is tempting to imagine automated help. However, it should be clear that the final allocation of an airline for its flights depends on the decisions made by all the other airlines. As such, unless there is dominating strategy for all the airlines, there is no easy way to create an automated mechanism. Automation can help in cost pre-computations for example. The likely most important indicator describing the goodness of the mechanisms is the fairness, or perceived fairness. BEACON explored fairness through the different gains, or relative gains, made by the airlines. However, it is not clear which particular fairness indicators should be used, i.e. which situation should be considered fair. We thus suggest any future research to tackle this problem early on, for instance trying to find a consensus among a large panel of airlines. In fact, we believe that this problem is bigger that the hotspot resolution issue, and that policymakers should clarify the indicator to be used for fairness in general in Europe after consultation with the airlines. The limitations and weaknesses of different types of agents (i.e. behavioural economics/gaming aspects) and cost approximations are described in detail in deliverables D5.2, D6.1, and D6.2.

TRL1.8	Do fundamental research results show contribution to the Programme strategic objectives e.g. performance ambitions identified at the ATM MP Level?	Achieved	The BEACON results mostly do not contribute directly to the performance ambitions, but they contribute to the transversal activities and betterment of the performance assessment methodologies. Furthermore, the choice of mechanisms and the modelling framework used to assess them has fruitfully shown several important issues that are generalisable and need to be addressed in future research related to OI AUO-0106 and OI AUO-0107 (see section 4 of D1.2)). However, the new mechanisms designed with the BEACON methodology may be the basis for more realistic future designs, and in this sense contributes to the Programme strategic objectives through these two OIs in particular.
TRL1.9	Have stakeholders been identified, consulted and involved in the assessment of the results?. Has their feedback been documented in project deliverables? Have stakeholders shown their interest on the proposed solution?	Achieved	Some of the stakeholders (like the airlines) have been identified, consulted and directly involved in the initial research design and the assessment of results. The stakeholder workshop at the beginning of the project collected the feedback on the initial BEACON mechanisms' design (see deliverable D6.1 for details). Furthermore, the stakeholders participated in the human-in-the-loop simulations, testing the mechanisms in the real-time and offering their feedback. These simulations were designed to obtain a further calibration of behavioural models and get the feedback on the mechanisms, but the latter motivation proved to be much more fruitful. Other stakeholders (like airports) have also been identified by the project, but not directly involved in the design of the mechanisms or in subsequent discussions. Some discussions are reported in D6.1, human-in-the-loop simulations can be found in D5.2 and some final considerations about stakeholders are included in D6.2.
TRL1.10	Have initial scientific observations been communicated and disseminated (e.g. technical reports/journals/conference papers)?	Achieved	Due to the multiple delays incurred by the project, only a few dissemination events have been attended by the BEACON team. However, the team has participated to a few conferences and submitted one conference paper so far, and has a list of events for after the project, as well as two draft articles to be submitted in journals in the next weeks. The final dissemination event is planned after the end of the project, to be able to disseminate the final results to a wider audience.

			The project has been selected for presentation at the Aviation World in Geneve, in March 2023, and several articles are almost ready for submissions in peer-reviewed journals. Furthermore, the updated values of cost of delay (D3.2) have been used in the modelling efforts of BEACON and NOSTROMO projects. The University of Westminster proposes to share this deliverable with the PRU and EUROCONTROL ('Standard Inputs for Economic Analyses') team, who currently use Westminster's cost of delay reporting [10].
TRL1.11	Are recommendations for further scientific research documented?	Achieved	Behavioural effects are numerous and only a small subset of the potential effects was tested in BEACON, and we suggest to continue this research to obtain a better assessment of these effects by having more extensive human-in-the-loop simulations and a more tailored survey. The gaming effects proved to be important to take into account. Any further research thus needs to find better ways to solve the rationality problem, potentially with reinforcement learning. We suggest the creation of a unique, standard dataset with regulation data in order to easily compare mechanisms tested by several projects. This dataset could be based on historical data but properly anonymised for public use. We suggest to explore the question of cost confidentiality through wider stakeholder consultations. Next avenue for the research lies in extending the slot swapping mechanisms to en-route airspace, thus having a more uniform prioritisation process across multiple time and geographical scales. The related issues of dynamicity and constraints from other stakeholders (like airports) should also be taken into account. The next steps and recommendation for further research work are documented in D6.2 and in the section 4.3 of D1.2.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

4 Conclusion and Lessons Learned

4.1 Conclusions

The project successfully designed and estimated the impact of several new mechanisms for hotspot resolutions, using detailed models including notions of behavioural economics.

First, the new mechanisms were designed by after intensive discussions about the current situation, the gains to be made, and the potential theoretical limitations. The feedback from the airspace users allowed us to explore good potential candidates for making new gains in hotspot resolutions mechanisms. The result is a balanced set of mechanisms, with a first one (ISTOP) showing the benefits of keeping some characteristics from UDPP (final choice to the airline, no airline loses from allocation etc.) while exchanging slots among airlines, which could be implemented in the mid-term future if needed. The second mechanism showed the benefits of a global optimisation process, with the introduction of credits as a balance to gaming effects, and also more exploratory, with serious modifications to the current operational process required in order to be implemented. The third one showed the benefits of letting the airline choose for themselves which slots are good for their flights and to which price (in credits), a typical market-driven feature. This mechanism is even more futuristic, since it would require a serious change of paradigm internal to the airlines, with purely (automated) cost-driven decisions from them[13].

The choice of mechanisms and the modelling framework used to assess them has fruitfully shown several important issues that are generalisable and need to be addressed in future research related to OI AUO-0106 and OI AUO-0107. In particular, the high potential of UDPP (at least when airlines know their true costs) and the degradation of information on the airline costs impacting the mechanisms should be taken into account in any new mechanism in this field. The important impact of gaming and behavioural effects is also to be taken account, and it is the firm belief of the consortium that no inter-airlines mechanism can be assessed properly without taking these effects into account. Note that a perfect estimation of the gaming effect was not achieved within the project, and will require more work (see bullet point about reinforcement learning in the next section). In addition, the first simulations with the small-scale model revealed that the true potential of flight prioritization mechanisms to reduce airline costs is only appreciated when regulations are large enough to involve delays potentially close to breaking operational margins for flights.

The results obtained on equity show that more effort is required in this matter. The fact that there is no consensus on the kind of equity that needs to be enforced, crossed with the relative inefficiency of some mechanisms like CM to ensure it, needs to be properly addressed. Indeed, equity is one of the key factors conditioning the acceptance of a new hotspot resolution mechanism by the airline. The equity, or fairness, or at least the ***perceived*** fairness of the impact of the mechanism will indeed play a major role in the final form of the mechanism. For instance, the fact that airlines may lose locally (i.e. in some regulations) from the mechanism may be hard to accept for them, even if the benefits in the long run (for them) are proved.

[13] Futuristic, but in line with other SESAR concepts, like 4D Trajectory Management.

Fairness is not, however, the only parameter that drives the potential adoption of new mechanisms. We flagged several of them during the projects, resulting from our discussions with airlines, which include the reluctance to use real money in operations, and, of course, the easiness to use the tool. This aspect is highly non-trivial, as it is linked to the question of automation within a competitive environment. Indeed, as shown during the project when designing the rational agents, the decisions made by the airlines during inter-airline mechanisms like CM or the auction are not independent from each other, on the contrary of UDPP. This means that airlines cannot just compute ***the*** best decision with the flick of a button[14]. We have shown that finding the best strategy (from the game theory point of view) is not obvious in general, which complexifies the decisions-making process and may render the mechanisms impractical in practice. More study about the heuristics that humans would use when faced with complex mechanisms (like the auction) is thus needed. The human-in-the-loop simulations were partly designed to answer to this question but only gathered anecdotal evidence.

Finally, we come back to one of the most important conclusions from the modelling process, related our first point on UDPP. Indeed, we would like to highlight the major issue of having approximated costs for subsequent optimisations. As highlighted in D5.1 in particular, this issue is central. If the airlines do not know their true costs, the mechanisms will likely be inefficient. The emphasis in this case should thus be on "helping" the airlines to find the best situation of them, which is the main like of research in UDPP *per se*. But even in a futuristic world where airlines would know their costs, the (necessary?) degradation of the information in some mechanisms would have the same impact. As a result, we suggest any future line of research in this area to tackle this issue and avoid assuming unrealistic degree of accuracy on the costs coming from the airlines. We also highlight the need to have a systematic assessment of the impact of the approximation, for instance with an extensive sensitivity analysis of the efficiency of the mechanisms to errors in the costs.

Regarding the specific objectives of the project, we present in Table 4 a summary of them and of their completion.

Table 4: Project objectives and their completion.

Objective	Completion
*"Propose a set of **improved flight prioritisation mechanisms** that expand current UDPP capabilities."*	Three new mechanisms have been proposed, who theoretically allows to have better cost reduction than UDPP, with incremental difficulty to implement: ISTOP, Credit Mechanism, Auction.
"*Define **new metrics** to evaluate the **fairness** and **equity** of flight prioritisation mechanisms and validate their appropriateness with AUs."*	Three indicators were defined based on previous work from ECTL on fairness and equity. They represent three distinct points of view on equity.
"Quantify the impact of 'non-rational' behaviours of AUs on the outcome of the proposed mechanisms, taking advantage of the	Gaming effects were estimated to trigger around 7 percentage points of reduction in economic

[14] Even if they know their true costs!

 Co-funded by the European Union

*methods and tools developed in the field of **behavioural economics**."*	efficiency and behavioural effects another 5 points of reduction.
*"Integrate the insights gained from behavioural economics into an **agent-based microsimulation** model of the full ECAC network able to capture network effects. The model shall be able to compute a set of key performance indicators (KPIs), including newly developed fairness and equity KPIs, allowing a comprehensive assessment of the new UDPP mechanisms."*	Two mechanisms were implemented in fast-time simulators fed with data from Mercury, a large scale passenger-based simulator. Data from 21 of most important European airports were used in the simulations, and KPIs related to Equity and Fairness were computed on top of standard PIs in delay and cost reduction efficiency.
*"Run a set of **simulation experiments** to evaluate the impact of the new UDPP mechanisms on the selected KPIs, taking into account behavioural effects, in order to analyse the advantages and the risks with respect to the current UDPP capabilities."*	We ran extensive experiments in different setups to estimate the efficiency and the equity of the new mechanisms, as well as the gaming and behavioural effects on the performance of the mechanisms.

4.2 Technical Lessons Learned

The lessons learned during BEACON are many, and we summarise the most important ones below:

- Gaming effects seem to be as strong as the behavioural ones in the results of WP5. As a result, one needs to take them into account carefully in the modelling process. Going beyond the simplified model that was used in BEACON, (multi-agent) reinforcement learning is probably the right tool to tackle this issue.
- The responses to the survey were very few, despite numerous efforts, several dissemination exercises, and three modifications of the introductory text to the survey. The exact reasons for this are unknown, even though we can assume that the COVID period, during which many airlines had some staff in hibernation, played a major role. We also assume that participants did not fully grasp the usefulness of the survey, since in appearance the questions were not strictly related to their job, or even to ATM[15]. For another future survey like this one, one may want to dedicate some effort to design a survey whose theme is more related to the participants' daily job. However, it must be highlighted that this is not an easy task and may require some intensive effort to find relevant questions that can still be used to estimate behavioural parameters.
- The human-in-the-loop simulations were mainly fruitful, the HMI and the backend working well in most cases. However, because of the COVID period, the consortium had no other choice than organising fully on-line sessions. This proved difficult, since the member of the team has

[15] As a reminder, the survey (which can be found in D5.2) was designed to catch risk aversion and other characteristics linked to behavioural economics, with questions that were not directly related to the ATM world.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

less control over some aspects of the experiments, for instance the browser used by the participants – which in some cases proved to be incompatible with the HMI – or some corporate firewalls blocking the connection. We thus suggest that these kinds of simulations, unless one wants to reach a mass audience, should be performed face-to-face.

- An important share of the effort related to the models went into the development of the Hotspot library. While this library now represents a major outcome of the project, that can be used in the future, its development was hindered by the lack of clear requirements from both models (from WP4 and WP5) and a missing list of the assumptions made by the initial version of the library. The lesson learnt here is thus to spend some effort initially in the documentation of an existing library before adapting it, especially if it is supposed to be used by distinct simulators with different assumptions and flows.
- While the plan for BEACON was to develop a first toy model followed by a more advanced one, we found that the project lacked an initial "trial" stage to test the mechanisms in extremely simple setups, fully controllable (for instance in terms of the size of regulations, size of airlines involved in the regulation etc.). While having a toy model and a more advanced one is very useful in general, we suggest having a bigger conceptual step between them, in order to have a more efficient exploration of the parameter space and the possible flavours of the mechanisms and agents first.
- Reinforcement learning (RL), flagged already as the right framework for a better computation of the gaming effect, may prove hard to solve in some cases. Some initial tests were performed during the project, which showed that the high degree of stochasticity of the system (in terms of size of regulation etc.) may prove a major barrier to solve the problem with RL. In line with the previous comment, we thus suggest using it in very controlled environment first, with low stochasticity in particular.

4.3 Plan for next R&D phase (Next steps)

Based on the conclusions and lessons learned, we suggest the following possible lines of work for the future of UDPP-related processes. The bullets are sorted by increasing applicated-oriented aspects

Fundamental research:

- The gaming effects, although not flagged as such at the beginning of the project, proved to be important to take into account. Any further research thus needs to find better ways to solve the rationality problem, potentially with reinforcement learning.
- Behavioural effects are numerous and only a small subset of the potential effects was tested in BEACON. It is likely that other effects like endowment plays a major role in the decision-making process of the airlines in the context of slot/flight allocations.
- More generally, and given the fact that some mechanisms seem to be sensitive to behavioural effects, we suggest to have a better assessment of these effects by having more extensive human-in-the-loop simulations and a more tailored survey.

Applied research:

- One of the benchmarks used during the project, NNBOUND, happened to have good characteristics both in terms of efficiency and equity. A potential line of research thus exists and would be to design a mechanism inspired by this benchmark, potentially merging it with CM or ISTOP.

- We suggest the creation of a unique, standard dataset with regulation data in order to easily compare mechanisms tested by several projects. This dataset could be based on historical data but properly anonymised for public use.

- BEACON only touched upon the problem of cost confidentiality, i.e. the fact that airlines may be reluctant to share their true costs, which is related to their business model. Other projects like SlotMachine[16] suggested the use of blockchain technology to solve this problem, while BEACON focused either on transforming the information (via a renormalisation process in ISTOP or bids in the Auction) and directly by assuming that the creation of an independent central planner that is trusted not to divulge this information is enough. We suggest going deeper in this direction, by consulting airlines and finding a consensus on the matter. This consensus may then drive the design of new mechanisms (for instance, not renormalising costs in ISTOP).

- Similarly, the consortium has noted a lack of consensus over the exact goal of potential new mechanisms. While the FPFS rule minimises the total delay, which apparently looks fair to airlines, one needs a new “objective” function related to airline, which could take into account equity by design. We thus suggest organising vast consultation activities to reach a consensus on this point (in short, setting KPIs for UDPP). This objective function could then also be used in other processes across the ATM system, like cost-driven extended arrival management.

- Finally, we note the crucial issue of a multi-sector hotspot resolution mechanism, and, more generally, of the additional constraints from other stakeholders. While this was flagged a potential issue to tackle during the first stage of the project, it was not explored. We thus suggest reflecting on the possibility to extend current mechanisms and exploratory ones to the ATM system in general, in particular to en-route airspaces, thus having a more uniform prioritisation process across multiple time and geographical scales. The related issues of dynamicity (e.g. change of conditions after a hotspot has been solved) and constraints from other stakeholders (like airports) should also be taken into account at the same time.

[16] SESAR 2020 project (http://slotmachine.frequentis.com/).

5 References

5.1 Project Deliverables

[1] BEACON project. D1.1 Project Management Plan, Edition 01.03.00, February 2021. (Not public.)

[2] BEACON project. D1.2 Final Project Results Report, Edition 01.00.00, 14th November 2022, approval pending.

[3] BEACON project. D2.1 Data Management Plan, Edition 01.00.01, February 2021. Available: https://www.beacon-sesar.eu/wp-content/uploads/2021/09/893100-BEACON-D2.1-Data-Management-Plan-V.01.00.01.pdf

[4] BEACON project. D2.2 Database structure and data elaboration, Edition 01.00.00, August 2022. Available: https://www.beacon-sesar.eu/wp-content/uploads/2022/10/893100-BEACON-D2.2-Database-structure-and-data-elaboration_v01.01.00.pdf

[5] BEACON project. D3.1 High-level modelling requirements, Edition 01.00.00, March 2021. Available: https://www.beacon-sesar.eu/wp-content/uploads/2021/09/893100-BEACON-D3.1-High-level-modelling-requirements-v01.00.00.pdf

[6] BEACON project. D3.2 Industry briefing on updates to the European cost of delay, Edition 01.01.00, November 2021. Available: https://www.beacon-sesar.eu/wp-content/uploads/2022/10/893100-BEACON-D3.2-Industry-briefing-on-updates-to-the-European-cost-of-delay-V.01.01.00-1.pdf

[7] BEACON project. D4.1 Behavioural Model Parameter Calibration, Edition 01.00.03, December 2021. Available: https://www.beacon-sesar.eu/wp-content/uploads/2022/10/893100-BEACON-D4.1-Behavioural-Model-Parameter-Calibration-V.01.00.03-1.pdf

[8] BEACON project. D4.2 Final model results, Edition 01.01.00, August 2022. Available: https://www.beacon-sesar.eu/wp-content/uploads/2022/10/893100-BEACON-D4.2-Final-model-results-V.01.00.00.pdf

[9] BEACON project. D5.1 First tactical model and results, Edition 01.01.00, November 2022. Available: https://www.beacon-sesar.eu/wp-content/uploads/2022/12/893100-BEACON-D5.1-First-tactical-model-and-results-V.01.01.00.pdf

[10] BEACON project. D5.2 Final tactical model and results, Edition 01.00.05, January 2023. Available here: https://www.beacon-sesar.eu/wp-content/uploads/2023/02/893100-BEACON-D5.2-Final-tactical-model-and-results-V.01.00.06.pdf

[11] BEACON project. D6.1 Intermediate concept assessment report, Edition 01.00.01, March 2021. Available: https://www.beacon-sesar.eu/wp-content/uploads/2021/09/893100-BEACON-D6.1-Intermediate-concept-assessment-report-V.01.00.01.pdf

[12] BEACON project. D6.2 Final concept assessment report, Edition 01.01.00, November 2022. Available here: https://www.beacon-sesar.eu/wp-content/uploads/2023/02/893100-BEACON-D6.2-Final-Concept-Assessment-report-V.01.02.00.pdf

[13] BEACON project. D7.1 Final dissemination and communication report, edition 01.00.02 January 2023. Available here: https://www.beacon-sesar.eu/wp-content/uploads/2023/02/893100_BEACON-D7.1-Final-dissemination-and-communication-report-V.01.00.02.pdf

[14] BEACON project. D8.1 H - Requirement No. 1, Edition 01.01.01, March 2021. (Not public.)

[15] BEACON project. D8.2 POPD - Requirement No. 5, Edition 01.01.01, March 2021. (Not public.)

[16] BEACON project. D8.3 POPD - Requirement No. 7, Edition 01.01.01, March 2021. (Not public.)

5.2 Project Publications

[1] BEACON project website. https://www.beacon-sesar.eu/.

[2] Poster for SID 2021, available on the SID website: https://www.sesarju.eu/node/3438

[3] Poster for SID 2020, available on the SID website: https://www.sesarju.eu/node/3438

[4] Presentation at the AGIFORS 2022, won the 'best innovation' paper award: https://publications.agifors.org/documents/1-04%20Andrea%20Gasparin%20ISTOP.pdf

[5] Presentation at the AGIFORS Airline Operations SG meeting 2021.

5.3 Other

[6] BEACON project. Grant Agreement 893100, Ref. Ares(2020)2203330 - 23/04/2020.

[7] Project Execution Guidelines for SESAR Exploratory Research, Edition 01.00.00, 08/02/2016

[8] European ATM Master Plan

[9] Gilovich, T., Griffin, D., and Kahneman, D. (Eds.), 2002. Heuristics and biases: The psychology of intuitive judgment. New York, NY, US: Cambridge University Press.

[10] University of Westminster (2015). European airline delay cost reference values Updated and extended values Version 4.1. Available at: https://www.eurocontrol.int/sites/default/files/publication/files/european-airline-delay-cost-reference-values-final-report-4-1.pdf

Appendix A

A.1 Acronyms and Terminology

Table 5: Acronyms and technology

Term	Definition
ANSP	Air navigation service provider
ATM	Air traffic management
AU	Airspace user
CM	Credit Mechanism
ECAC	European Civil Aviation Conference
EN	enabler
ETA	Estimated Time of Arrival
EUFALDA	European Federation of Airline Dispatchers Association
FPFS	First Planned First Served
HITL	Human-in-the-loop
HMI	Human-machine interface
IATA	International Air Transport Association
ISTOP	Inter-airline Slot Trading Offer Provider
KPI	Key performance indicator
NM	Network Manager
OI	Operational improvement
PRU	Performance Review Unit
RL	Reinforcement learning
S3JU	SESAR3 Joint Undertaking (Agency of the European Commission)
SESAR	Single European Sky ATM Research Programme
SFP	Selective Flight Protection
UDPP	User Driven Prioritisation Process

Appendix B Experimental plan

See next pages.

BEACON experimental plan draft

Deliverable ID:	NA
Dissemination Level:	CO
Project Acronym:	BEACON
Grant:	893100
Call:	H2020-SESAR-2019-2
Topic:	SESAR-ER4-08-2019
Consortium Coordinator:	University of Westminster
Edition date:	10/11/2022
Edition:	01.01.00
Template Edition:	02.00.03

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Authoring & Approval

The beneficiaries/consortium confirm(s) the correct application of the GA, which includes data protection provisions, and compliance with GDPR or the applicable legal framework with an equivalent level of protection, in the frame of the Action. In particular, beneficiaries/consortium confirm(s) to be up to date with their consent management system.

Authors of the document

Name / Beneficiary	Position / Title	Date
Gérald Gurtner / UoW	Project coordinator	10/11/2022

Reviewers internal to the project

Name / Beneficiary	Position / Title	Date
Tatjana Bolic / UoW	Project beneficiary	28/05/2022

Reviewers external to the project

Name / Beneficiary	Position / Title	Date

Approved for submission to the SJU By - Representatives of all beneficiaries involved in the project

Name / Beneficiary	Position / Title	Date
Gérald Gurtner	Project coordinator	10/11/2022

Rejected By - Representatives of beneficiaries involved in the project

Name and/or Beneficiary	Position / Title	Date

Document History

Edition	Date	Status	Name / Beneficiary	Justification
00.00.01	20/05/2022	Draft	Gérald Gurtner	Initial draft
00.00.02	28/05/2022	Draft	Gérald Gurtner	Second draft
01.00.00	31/05/2022	Rejected	Gérald Gurtner	For SJU submission
01.01.00	10/11/2022	Submitted	Gérald Gurtner	For SJU second submission

EUROPEAN PARTNERSHIP

Co-funded by the European Union

BEACON

BEHAVIOURAL ECONOMICS FOR ATM CONCEPTS

This technical report is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 893100 under European Union's Horizon 2020 research and innovation programme.

Abstract

This document describes the experimental plan for BEACON, gathering information from several deliverables describing the experiments planned or carried out. The plan is structured as to follow the ER experimental guidance document published by the SJU for ER project.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Table of Contents

List of Tables

List of Figures

1 Introduction

BEACON's general goal is to design new procedures for airspace users (AUs) to better allocate their resources (aircraft, pilots, crew, and others) in case of disruptions and evaluate the proposed procedures through new methods and tools able to take into account AUs' complex behaviours, such as bounded rationality.

More specifically, BEACON has the following high-level objectives:

1. Propose a set of improved flight prioritisation mechanisms that expand current user driven prioritisation process (UDPP) capabilities.

2. Define new metrics to evaluate the fairness and equity of flight prioritisation mechanisms and validate their appropriateness with AUs.

3. Quantify the impact of 'non-rational' behaviours of AUs on the outcome of the proposed mechanisms, taking advantage of the methods and tools developed in the field of behavioural economics.

4. Integrate the insights gained from behavioural economics into an agent-based microsimulation model of the full ECAC network able to capture network effects. The model shall be able to compute a set of key performance indicators (KPIs), including the newly developed fairness and equity KPIs, allowing a comprehensive assessment of the new UDPP mechanisms.

5. Run a set of simulation experiments to evaluate the impact of the new UDPP mechanisms on the selected KPIs, taking into account behavioural effects, in order to analyse the advantages and the risks with respect to the current UDPP capabilities.

6. Derive guidelines and methodological recommendations on the further development, validation, and deployment of the new UDPP mechanisms that pave the way to a more harmonised and efficient flight prioritisation process across Europe.

Several experiments were planned in order to reach these objectives. First, the project planned to prepare a questionnaire, to collect data to calibrate different behavioural effects in the subsequent models. Second, human-in-the-loop simulations were planned in order to gain insight into real human decision-making processes, in particular, again, in relation to behavioural effects. Third, fast time simulations were planned too, with two different models. The first model, developed in WP4, was planned to deliver first insights into the inner working of mechanisms and a first assessment of the impact of behavioural effects on the mechanisms. The second model, developed in WP5, was planned to do some large-scale network simulations with a more detailed model in order to have more realistic KPI assessments.

This experimental plan summarises the experimental approaches for each of these, using as a guideline the ER experimental guidance [1] and summarising the explanations already contained, or planned to be contained in technical deliverables.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

2 Overview of experimental approach

2.1 Questionnaire

2.1.1 Research questions

The questionnaire was planned in order to collect data to calibrate some behavioural effects in the subsequent models. Indeed, the behavioural frameworks we use in BEACON, prospect theory and hyperbolic discounting, use several parameters that need to be chosen.

2.1.2 Experiment contribution to project objectives

The answers to the questions specifically chosen for the questionnaire, reveal how humans behave in certain situations that may elicit non-rational behaviours. By studying the answers, the project would be able to compute the parameters needed for the calibration of prospect theory and hyperbolic discounting rules.

2.1.3 Experimental setup

The questionnaire is designed by using typical situations in which behavioural effects captured by hyperbolic discounting and prospect can appear. Typically, participants are asked to answer questions pertaining to losses and gains, e.g:

“Would you rather play to lottery A or B?

- A: you have 50% or winning 50€ and 50% chance of losing 20€
- B: you have 50% chance of winning 10000€ and 50% chance of losing 9980€.”

The questionnaire was prepared online, with a minimal introduction to the aim of the study for the participants. Indeed, giving them too much information may influence their answers. The questionnaire is included in the appendix of D4.1 [9].

The answers are then supposed to be analysed using regressions to estimate statistically the relevant parameters.

See also more details about the consent and ethical aspects of the experiments in deliverables D8.X [15][16][17].

2.1.4 Planned experiments

The recruitment of participants was done using various means, targeting the population that would know enough about flight dispatching/fleet management/ATFM regulation. Indeed, it is known that the answers to such questionnaires can be quite different depending on the profile of the respondent, including gender, working position etc.

We had two major pushes with the survey, over a long period of time. In the first attempt, the survey went online on 30th April 2021, and was shared with European Federation of Airline Dispatchers

Association (EUFALDA), as the EUFALDA president agreed to share the link to the survey with their members. The survey was further publicised through BEACON website and BEACON Linked-in account and shared directly via consortium and advisory board members. Five months later we had one response.

The second attempt was agreed with IATA. The consortium members met with IATA representatives to discuss how they could help with the survey. The consortium proposed to organise a short workshop with dispatchers to explain the reasons for collecting the data and why the survey does not really touch on aviation questions, and then to ask the participants to fill in the survey. We were advised to share the survey link with IATA, as due to the high uncertainty under which the airlines were operating, it was close to impossible to get even an hour or two from dispatchers that are not linked directly to their work. Thus, IATA agreed to send the survey to their members, suggesting to fill it in. This second push started on 3rd December 2021. In March 2022 we obtained two additional responses.

In addition to these two major pushes, ECTL also contacted the advisory board and UDPP contacts in three separate occasions (15/07/21, 31/08/21 and 15/12/21) to push them to take the survey.

2.1.5 Results expected

The results expected were the parameters regressed on the answers to the questionnaire. Unfortunately, the project has had very few responses (3 complete responses in total), which prevented the calibration of parameters entering the functional forms of prospect theory and hyperbolic discounting (see D4.1) using the questionnaire-collected data, and used instead parameters values found during a literature review.

2.2 Human-in-the-loop simulations

2.2.1 Research questions

Behavioural effects are notoriously difficult to catch. Not because they have little effect on humans' decisions, but because they are hard to isolate from each other. The questionnaire aimed at catching some of them in a clean and controlled way. However, the context in which they were supposed to be captured was quite far from the actual decision-making process made by humans during the resolution of ATFM regulations (see sample question in section 2.1.3).

Human-in-the-loop simulations are meant to do the opposite: have a more realistic environment, but focus less on a specific situation. This more holistic point of view allows to have more qualitative insight into the decision-making process but in some cases may also be used for regression and statistical analyses.

In BEACON, these were exactly the goals: get feedback on the mechanisms from participants and, if possible, compare the previous calibration (from the questionnaire) to the data collected in the human-in-the-loop simulations. On top of that, these simulations allow to have an assessment of the efficiency (as well as other metrics) of the mechanism itself with real decision-making processes, i.e. from humans, as opposed to artificial agents.

2.2.2 Experiment contribution to project objectives

EUROPEAN PARTNERSHIP

Co-funded by the European Union

When it comes to assessing behavioural effects, nothing compares to experiments with real humans. These experiments help us understand how the humans behave, which is important not only to get knowledge on their behaviour, but also to be able to replicate it, to some extent, using the framework from behavioural science.

Finally, the feedback from the participants is important for the mechanisms themselves, how they are currently designed and how they could evolve in the future for instance in terms of vocabulary, semantics, and visuals, in order to be used in the ATM system.

2.2.3 Experimental setup

The model developed in WP5 (Mercury) was chosen to serve as the backend for the simulations. An interface enabling interaction of participants with the simulator was developed by NOM. Furthermore, a server to enable communication between the interface and Mercury was developed by UoW. Thus the data coming from Mercury was displayed to the participants via interface, which was also collecting prioritisation data from participants and sending them back to Mercury (via server). After receiving the information from the participant, the Mercury would compute the resolution of the regulations. This process is summarised in Figure 1.

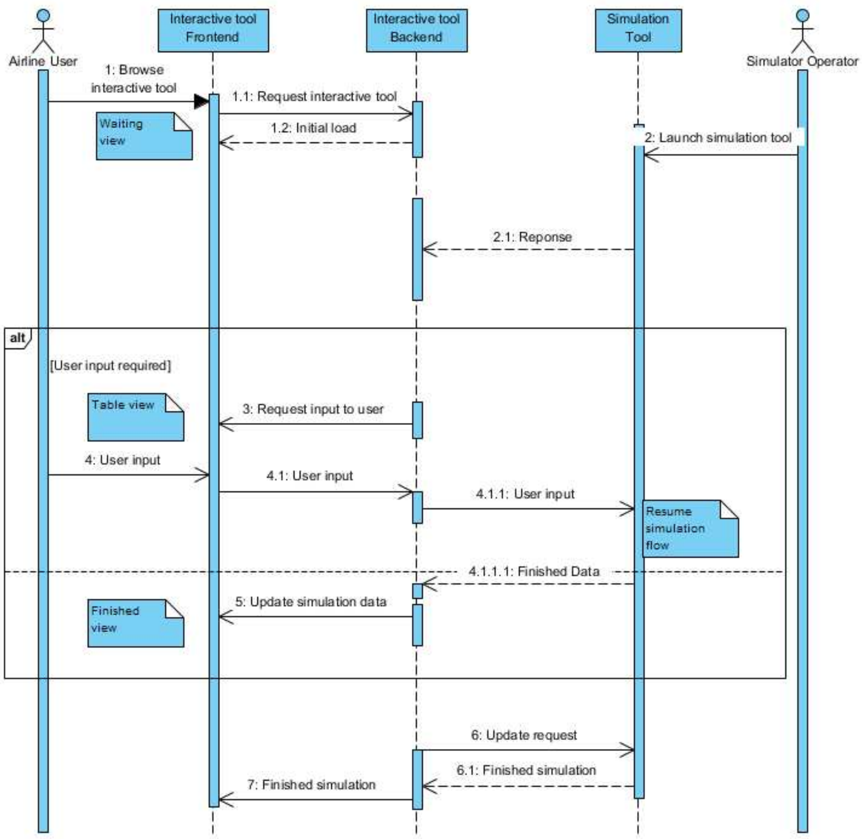

Figure 1: communication flow diagram between user, server, HMI, and server operator.

It was decided, to focus on regulations happening at a specific airport (CDG). This airport was chosen for its considerable size while having other desirable properties (like regulation applied to its TMA directly, as opposed to neighbouring sectors like London-based airports). The participant played a specific airline, Air France -like, chosen for its number (i.e. large) of flights landing at CDG.

In short, the model was starting a simulation of flights and passengers moving through Europe, until a regulation hit CDG. At this point, information on the flights the participant was supposed to manage were sent to the interface. The player had to make several decisions concerning these flights (e.g. prioritise them). The product of these decisions was sent back to the model, which, with the decisions coming from other artificial agents representing the other airlines involved in the regulation, were used to solve the regulation. The final slot allocation was then sent back to the player, along with other relevant information, serving as feedback for their decisions.

2.2.4 Planned experiments

Two different mechanisms were tested with each participant (credit mechanism and ISTOP). The order in which these mechanisms appeared to the player was randomised. At the end of the session, 10 minutes of discussion were dedicated to feedback, with a set of fixed questions asked every time, plus some general feedback from the player about the mechanisms.

The participants were required to have an advanced knowledge of flight dispatching, ideally being flight dispatchers or having a similar role in an airline. Seven participants were recruited in total, all of them coming from airlines directly or having very close relationships with them. Introductory slides were presented to the participants beforehand, explaining the aim of the study, mechanisms, the interface and the role that they take during the experiments. Each participant played each mechanism for about and hour. The BEACON team was always in contact with participants during the simulations, to help them with technical issues. The experiments took place from end of March to end of May, to suit the schedule of the participants.

2.2.5 Results expected

The results expected are of three types:

- A further calibration of the behavioural parameters, if possible.
- General feedback, possibly related to the mechanisms themselves, or pertaining to non-rationality (but anecdotal).
- Efficiency of the mechanism, estimated with real humans, as opposed to artificial agents in the fast-time simulations.

2.3 Fast-time simulations in WP4, small scale

2.3.1 Research questions

New trading or optimisation mechanisms have many possible flavours and alternatives. In order to test them and understand the main advantages and issues of each of them, it is important to have a model which is fairly simple while retaining the main characteristics of the system under scrutiny. The model built in WP4 is aiming exactly at doing that – allow mechanisms to be applied in the small-scale synthetic network on the realistic, albeit synthetic, traffic and airline data.

Moreover, the behavioural effects are not always easy to conceptualise and implement in new situations. Their effects might be compounded with other ones – like network effects, and it is thus more appropriate to test them in a small-scale model where different effects are more clearly seen in isolation.

More information on the WP4 model, as well as the results of the experiments, can be found in D4.2 [10].

2.3.2 Experiment contribution to project objectives

The experiments in WP4, quick and easy thanks to the small scale of the model, allow to have some first insights on the mechanisms, which are brand new, and the behavioural effects emerging from airline agents' actions. The average efficiency and the effects of the deviations from rationality in each of them can be quicky assessed with the model, leading to more informed implementation and focus in WP5.

2.3.3 Experimental setup

The model is built on DDR data. Five airports are selected in the dataset (2019, see D2.2 [6] and D4.2 [10] for more details on the dataset). All flights between these airports are selected, flying between 9.00 and 23.00. A synthetic airspace is built around the airports (as shown in D4.2). Rotations and passenger itineraries are built synthetically, airport and airspace capacities are adjusted, and regulations are created randomly throughout the airspace. Exogeneous delays are generated in order to stress the system and have more regulations.

The BEACON mechanisms are implemented, and airline agents "play" the mechanisms when regulations hit. Due to randomness, a certain number of simulations is performed on each setup, leading to a probabilistic analysis of the results.

The metrics collected include those on efficiency and fairness, which allows for a first approximated overview of the mechanisms with behavioural effects.

2.3.4 Planned experiments

Different experiments were planned for WP4, aiming at testing the mechanisms and behavioural effects. More specifically, experiments were run based:

- On different mechanisms implemented, one at a time: credit mechanism, auction, and ISTOP.

- On different behavioural effects, applier together or one at a time (for baseline): prospect theory and hyperbolic discounting.

Note that all experiments are compared with a baseline featuring the "FPFS mechanism" (First Plan First Served) without deviations form rationality.

2.3.5 Results expected

The results expected out of the model are the following:

- Computation of efficiency, fairness, and potentially other metrics for different mechanisms.
- Computation of efficiency, fairness, and potentially other metrics for different behavioural effects.

The original plan was that these insights would inform the implementation of the second model in WP5. Given the extra delays in WP4, only preliminary results from WP4 have been taken into account to develop WP5, selected the most interesting mechanisms.

2.4 Fast-time simulations in WP5

2.4.1 Research questions

The model in WP5 is meant to give a more accurate picture of the impact of the new mechanisms on the system, taking into account behavioural effects. The model developed there is thus more detailed and have a larger scope, capturing both more subtle behaviours and more large-scale effects. The performance indicators collected by running the model are expected to be more precise and more robust than those from WP4.

Due to the higher complexity of the model, WP4 goal was also to inform the decisions made in WP5, in order to avoid less interesting experiments being carried out. Hence, a selection of the most interesting mechanisms was performed, which will be included in the WP5 model.

2.4.2 Experiment contribution to project objectives

The model in WP5 is highly detailed and, while representing an idealised version of the air transportation system, allows to compute key performance indicators with a high quality. The fact various metrics are computable allow for various trade-off analyses and a holistic assessment of the mechanisms. Moreover, the detailed processes in the model allow to embed complex behaviours, such as the ones that we want to test in BEACON.

Hence, these fast-time simulations allow to assess the concepts of the new mechanisms and the effects of airlines' behaviour on them.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

2.4.3 Experimental setup

The experiments are based a vast array of data coming from various sources, including DDR data for flight plans and GDS for schedules and passenger itineraries (see D2.1 [5] for details). Data are cleaned and merged, a typical data of operation is selected, and a dataset is built for the model. The model includes various agents, including airline operating centres that will solve regulations when they hit, following different mechanisms. Regulations are also sampled from historical data and applied at random. The decisions taken by the airline operating centres are based on advanced cost model, taking into account curfew, passenger costs, and flight costs, among others.

Exogeneous delays (on flying times, passenger transfer, turnaround etc) are applied stochastically, based on probability distribution calibrated on data. Different mechanisms can be applied in case of regulations, replacing the default UDPP or FPFS rules.

Due to stochasticity, several iterations of the same simulation are performed, leading to an analysis taking into account the heterogeneity of airlines, flights, airports, etc. For Mercury, the typical number of iterations used is 100, giving a relative standard error below 5% on most indicators.

2.4.4 Planned experiments

The experiments planned mirror the ones in WP4:

- Different mechanisms implemented, one at a time.
- Different behavioural effects, applier together or one at a time (for baseline): prospect theory and hyperbolic discounting.

On top of that, Mercury will play with the scope of the simulation. Simulations with only one airport (CDG), similarly to the human-in-the-loop ones, will be performed for the first deliverable. The second deliverable will include larger scale simulations, trying to highlight the potential role of network effects. Moreover, depending on the initial tests, Mercury will also be used in different operational environments, stressed (high delays) or nominal (historical delays). It is obviously expected that the mechanisms will play a more important role in the stressed scenario, however the team is also interested in assessing how much can be gained in more normal situation.

2.4.5 Results expected

The output from Mercury contains a vast number of metrics. The project will focus on efficiency indicators for the mechanisms, derived from these metrics, as well as fairness and any other indicator that may be of interest. Trade-off analysis and advanced comparisons between scenarios, in particular in order to see the effects of behavioural effects, will be carried out.

3 Detailed experimental approach

The questions below are taken directly from [1].

3.1 Scientific approach

What is/are your (null/research) hypothesis/hypotheses? (if applicable)

- Respective mechanisms modify the baseline UDPP efficiency, equity, etc.
- BE has an effect on the efficiency and other KPIs for each mechanism.

Are there dependent, independent and control variables in your research, and what are they?

- Dependent variables will be documented in D4.2 [10] and D5.1 [11] for each respective model.
- Independent variables are described in the scenarios, see D3.1 [7].

Do you have different (validation) scenarios? Is there a reference scenario? Is there a solution scenario? What are the differences between these scenarios?

- Reference scenario is a typical day of September 2018 (2019 for WP4) with standard UDPP implemented (without BE).
- Solution scenarios have been built using a combination of traffic conditions, active mechanism and active biases.

What statistical tests will you be applying? What is their significance level and how will they be set? What software will you use for statistical evaluation?

- Survey results are analysed for instance with t-tests. A regression is performed using the PT functional form.
- Simulation results are typically analysed using Fisher tests, F-tests, or KS tests, with p-value adjustments (Type 2 error or Bonferroni).
-

Will you define and develop new Performance Indicators or adopting existing KPIs?

- We will use a blend of existing and new KPIs in BEACON.
- New KPIs are based on previous work and submitted to the AB feedback. They are documented in D3.1 [7].

3.2 Fast-time simulations

Are the planned simulations stochastic or deterministic? Will you model uncertainties, disturbances and disruptions? How?

- Both models in WP4 and WP5 are stochastic.
- Disturbances and disruptions happen either endogenously (for instance, too many flights trigger a regulation) or exogenously (a reduction of capacity is applied, modelling an external events).
- All disturbances and noise characteristics are based on historical data.

3.3 Questionnaire

Are you using validated questionnaires/scales or designing your own questionnaire?

- We designed our own questionnaire, adapting typical questionnaires and scales used in the field, to the UDPP addressed in BEACON.

Do you plan to use open-ended responses? How will you analyse them?

- We only used one open-ended question, relating to the position of the respondent.

Did you plan appropriate statistical tests determined by the scalar format of the question (e.g. categorical / ordinal / interval; dependent / independent; even vs. odd number of items on Likert scale)?

- The statistical tests applied are t-tests and other related tests, appropriate to the desired sample size.

Have you tested the questionnaire (in case you are designing your own questionnaire)?

- The questionnaire has been tested internally with individuals from the consortium.

3.4 Human in the loop simulations

What are the potential sequence effects (such as fatigue, learning, carry-over, maturation, reactivity) and how will your experimental/factorial design ensure you avoid them (e.g. Latin square, randomization, pseudo-randomization)?

- Pseudo-randomisation of scenarios is used.

If you are running an experiment, how many test subjects are you planning to include / how big is the sample? How do you estimate the required sample size?

- Only a handful participants participate in the simulations.
- Each participant participates in several runs.
- The exact number of runs is chosen based on the total time available from participants.

Which specific participants/respondents (demographics, expertise, characteristics) do you need?

- We are targeting fleet dispatchers or equivalent.

How will you recruit/sample the participants/respondents?

- Via direct contact with our AB and SWISS.

3.5 Data needs

What are your data needs? What type of data sets do you need for your research? (e.g. flight plan data, radar tracks, delay data, cost data, passenger itineraries, validated schedule data)

- Itineraries/schedules for 2018.
- Various data for cost of delay update.
- DDR data.

How will you obtain these data sets?

- We bought schedules and itineraries. Other data are acquired through our contacts in the industry.

How will you process/clean these data sets?

- Cleaning process includes removal of obvious mistakes.
- We then merge schedules and DDR, and then pax and schedules.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Will you use open data sources?

- Only anecdotally.

Do you have sufficient finances for data procurement? If insufficient, what would be your alternatives?

- We have enough budget for the data acquisition.

3.6 Dissemination

The consortium should make sure that the research data management has been planned (e.g. Storage, Security, and Access during data lifecycle).

- See D2.1 Data Management Plan [5].

If you plan to disseminate results by means of a scientific publication, do you also plan to make the underlying research data available? This is of particular relevance if the project participates to the H2020 open research data pilot.

- No input data can be shared. Output data may be shared after a period of embargo.

What would be necessary to reproduce the results?

- Description of applied methodologies (simulators) as well as data sources.

Would another researcher be able to reproduce them? Do you plan to make your source code open to facilitate reproducibility?

- Other researchers can re-implement the models we are using based on their description included in the deliverables. We do not plan to release the code of the model as open-source.

Are data, methodology descriptions, software available and public for this purpose? If not, why not?

- All software we are using for the models is freely available, in general in open source.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

4 References

[1] [SESAR 2020 General Document Template ER], edition 00.01.00, Dec. 2022
[2] Grant Agreement No 893100 BEACON – Annex 1 Description of the Action.
[3] BEACON Consortium Agreement.
[4] BEACON D1.1 Project Management Plan, Issue 1, July 2020.
[5] BEACON D2.1 Data Management Plan, Feb. 2021
[6] BEACON D2.2 Database structure and data elaboration, to be submitted.
[7] BEACON D3.1 High-level modelling requirements, May 2021.
[8] BEACON D3.2 Industry briefing on updates to the European cost of delay, Nov. 2021.
[9] BEACON D4.1 First model version and behavioural calibration, Dec. 2021.
[10] BEACON D4.2 Final model results, submitted May 2022.
[11] BEACON D5.1 First tactical model and results, to be submitted.
[12] BEACON D5.2 Final tactical model and results, to be submitted.
[13] BEACON D6.1 Intermediate Concept assessment report, May 2021.
[14] BEACON D6.2 Final Concept assessment report, to be submitted.
[15] BEACON D8.1 H - Requirement No. 1, May 2021.
[16] BEACON D8.2 POPD - Requirement No. 5, May 2021.
[17] BEACON D8.3 POPD - Requirement No. 7, May 2021.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Appendix A : Acronyms

Acronym	Meaning
ABM	Agent Based Model
ATM	Air Traffic Management
ATFM	Air Traffic Flow and Capacity Management
AU	Airspace User
ECAC	European Civil Aviation Conference
FPFS	First Planned First Served
HD	Hyperbolic Discounting
ISTOP	Inter Airline Slot Swap Offer Provider
KPA	Key Performance Area
KPI	Key Performance Indicator
PT	Prospect Theory
SESAR	Single European Sky ATM Research
SES	Single European Sky
UDPP	User Driven Prioritisation Process
WP	Work Package

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Appendix B : Summary

Table 1 summarises the experimental approach, taking topics and questions directly from [10] and referencing relevant documents.

Table 1: Summary of experimental and references

Overview aspects	Project approach	References
Definition of research questions	• Research questions have been explained in several documents, starting from the proposal • Goals of BEACON relate to core research question about UDPP and behavioural economics (BE)	• Grant Agreement • Presentation SESAR Innovation Days 2020 • Deliverables D3.1
Experiment contribution to project objectives	Several sets of experiments were planned carefully by the project to understand new UDPP mechanisms and BE: • A survey, notably to collect data for BE parameters calibration from the answers to the survey • Several human-in-the-loop simulations (HITL), for further calibration of BE parameters • Several preliminary simulated experiments with a small-scale model • Final simulations with a large-scale model, to compute KPIs more realistically	• Grant Agreement for rationale • Deliverable D3.1 for general context and the description of the assumptions and the scenarios • Deliverable D4.1 for the survey • Future deliverable D5.1 for the HITL simulations and the second model. • Deliverable D4.2 for the first model
Basic set up and methodology of the research	• Design new mechanisms for UDPP • Design scenarios where new mechanisms can be tested • Build a first simulator for a quick estimate of the impact of the mechanisms	• Design of scenarios, new mechanism, indicators, etc., all in D3.1. • Survey described in D4.1

EUROPEAN PARTNERSHIP

Co-funded by the European Union

	• Update an existing second simulator to estimate more realistically the impact of the mechanisms • Disseminate a survey to estimate behavioural parameters • Organise HITL simulations to better estimate BE parameters and efficiency of mechanisms with actual actors	• HITL simulations and second model described in D5.1 • First model described in D4.2
Planned experiments and interlinkage	• Survey feeds model A and model B • HITL simulation feeds model B and allows to compute network-level KPIs • Model A results on mechanisms feeds model B	• Same as above
Will your results be qualitative or quantitative (or both)?	• Most results of experiments are quantitative • Small qualitative insight provided by the survey	• See D3.1 regarding the indicators/quantitative aspects • See D4.1 for the qualitative aspects
How will you validate your outcome (concept, system, tool, model, scientific finding, etc.)?	• Survey and HITL simulations rely mostly on statistical analyses • Model validation relies on comparison with known situation and internal logic validation • AB and workshops help validating the feasibility of mechanisms	• See D6.1 regarding the concept assessment • See D4.1 for survey analysis • See deliverables D4.2, D5.1, and D5.2 for validation tests
Where is your experimental/exploration/validation/simulation exercise	• See thereafter	• Survey exploration and analysis is described in D4.1

planning documented?		• Human-in-the-loop simulation and survey plan can be found in Deliverables D8.X. • Scenarios, mechanisms and their testing can be found in D3.1. • Analyses will be reported in D4.2, D5.1, and D5.2 • The overall validation and concept assessment will be reported in D6.1
Overview aspects	• BEACON approach	• References
Will you try to measure your concept impact with respect to several potentially competing Performance Areas? Will your research help improving the understanding of the associated trade-offs?	• Yes, several KPAs will be covered (notably equity and punctuality) • We perform routine trade-off analyses out of our models, this will be the case for BEACON	• Indicators and their respective areas can be found in D3.1 • Trade-off analyses will be performed in D5.1 and D5.2.