



**HAL**  
open science

# Hero.coli: a video game empowering stealth learning of synthetic biology: a continuous analytics-driven game design approach

Raphaël Goujet

► **To cite this version:**

Raphaël Goujet. Hero.coli: a video game empowering stealth learning of synthetic biology: a continuous analytics-driven game design approach. Education. Université Sorbonne Paris Cité, 2018. English. NNT: 2018USPCB175 . tel-02524484

**HAL Id: tel-02524484**

**<https://theses.hal.science/tel-02524484>**

Submitted on 30 Mar 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ  
PARIS  
DESCARTES

U<sup>S</sup>PC  
Université Sorbonne  
Paris Cité

UNIVERSITÉ PARIS DESCARTES

Frontières du Vivant Doctoral School 474 - New Frontiers PhD  
Program

*Inserm U1001 - Center for Research and Interdisciplinarity*

# Hero.coli: a video game empowering stealth learning of synthetic biology

*a continuous analytics-driven game design approach*

By Raphaël Goujet

Interdisciplinary biology PhD thesis

Directed by Ariel Lindner

Publicly presented and defended on 30 November 2018

In front of a jury comprising:

Mr Sébastien GEORGE	Reviewer	Le Mans Université
Mr Jean-Marc LABAT	Reviewer	Université Pierre et Marie Curie
Mrs Patricia MARZIN-JANVIER	Examiner	Université de Bretagne Occidentale
Mrs Melanie STEGMAN	Examiner	Molecular Jig Games, LLC
Mr Ariel LINDNER	Supervisor	Université Paris-Descartes

 Except where otherwise noted, this work is licensed under <http://creativecommons.org/licenses/by-nc-nd/3.0/>

# Contents

<b>Contents</b>	<b>3</b>
<b>List of Figures</b>	<b>7</b>
<b>Synopsis</b>	<b>11</b>
<b>I Introduction</b>	<b>15</b>
<b>1 Video games and Learning</b>	<b>17</b>
1.1 Generalities about video games . . . . .	17
1.1.1 Game . . . . .	17
1.1.2 Gameplay and Gameplay Loops . . . . .	19
1.2 Digital game-based learning . . . . .	22
1.2.1 The need for game-based learning . . . . .	22
1.2.2 Applications of digital game-based learning . . . . .	23
1.2.3 Other Serious Games and GWAPs . . . . .	28
1.2.4 Closely-related genres . . . . .	30
1.3 Learning strategies, assessment, and outcomes . . . . .	34
1.3.1 Learning strategies . . . . .	34
1.3.2 Learning assessment . . . . .	34
1.3.3 Learning outcomes and issues . . . . .	36
<b>2 Synthetic biology</b>	<b>39</b>
2.1 Definition . . . . .	39
2.1.1 The Central Dogma . . . . .	40
2.2 Principles . . . . .	42
2.2.1 Decoupling . . . . .	42
2.2.2 Standardization . . . . .	42
2.2.3 Abstraction . . . . .	47
2.3 Limitations of synthetic biology . . . . .	56
2.3.1 Complexity . . . . .	56
2.3.2 Variation . . . . .	56
2.3.3 Evolution . . . . .	57
2.4 Uses of synthetic biology . . . . .	57

2.5	Dissemination . . . . .	59
2.5.1	Synthetic biology in popular culture . . . . .	60
2.5.2	Synthetic biology in academic training . . . . .	64
2.5.3	Synthetic biology in popular science on the Internet . . . . .	65
<b>3</b>	<b>Questions, approaches, and objectives</b>	<b>71</b>
3.1	Research questions . . . . .	71
3.2	Outlining the remaining chapters . . . . .	73
<b>II</b>	<b>Experimental Setup</b>	<b>75</b>
<b>4</b>	<b>Design and Implementation of Hero.Coli</b>	<b>77</b>
4.1	Genesis . . . . .	77
4.1.1	History of the CRI . . . . .	77
4.1.2	Synthetic biology . . . . .	78
4.1.3	Education . . . . .	78
4.1.4	The Digital Synthetic Biology Club . . . . .	78
4.1.5	Citizen Cyberlab . . . . .	79
4.2	Hero.Coli 1.12: a proof-of-concept . . . . .	81
4.2.1	First design of Hero.Coli . . . . .	81
4.2.2	Simulator . . . . .	92
4.2.3	Technical implementation . . . . .	93
4.2.4	Playtesting and accolades . . . . .	96
4.3	Repurposing into a new research project: Adaptations . . . . .	99
4.3.1	Accessibility . . . . .	99
4.3.2	Metrics and analytics . . . . .	100
4.3.3	Academic use . . . . .	103
4.3.4	Game design . . . . .	104
4.3.5	Tutorial . . . . .	105
4.3.6	Interface . . . . .	106
4.3.7	Simulator . . . . .	107
<b>III</b>	<b>Data gathering and analysis</b>	<b>109</b>
<b>5</b>	<b>Data gathering campaigns</b>	<b>111</b>
5.1	Survey methodology . . . . .	111
5.2	Online experiments . . . . .	114
5.3	In-class experiments . . . . .	115
5.4	Final 2018 experiment at Cité des Sciences . . . . .	119
5.4.1	Objectives . . . . .	120
5.4.2	Protocol . . . . .	120
5.4.3	Implementation . . . . .	121

<b>6</b>	<b>Data analysis</b>	<b>123</b>
6.1	Exploratory analysis . . . . .	123
6.1.1	Feedbacks from the participants . . . . .	123
6.1.2	Filtering the data . . . . .	128
6.1.3	Correlations in survey answers . . . . .	129
6.1.4	Correlations between tracking data and surveys . . . . .	132
6.1.5	Surveys: Analysis of the cohort . . . . .	136
6.2	Detailed analysis . . . . .	139
6.2.1	Threshold effect: learning and checkpoints . . . . .	139
6.2.2	Comparison of the pretest and posttest pairs . . . . .	143
6.2.3	Surveys and game metrics: data mining . . . . .	155
6.2.4	Comparison of phase 1 and phase 2 . . . . .	156
6.2.5	Limitations . . . . .	158
<b>7</b>	<b>Conclusions and prospectives</b>	<b>159</b>
7.1	Conclusions . . . . .	159
7.1.1	Research questions . . . . .	159
7.1.2	Usefulness, Usability, Acceptability . . . . .	162
7.2	Prospectives . . . . .	162
<b>A</b>	<b>Annex 1: tables</b>	<b>179</b>
<b>B</b>	<b>Annex 2: graphs</b>	<b>185</b>
B.1	Figures referenced in section 6.2.2 . . . . .	186
B.2	Figures referenced in sections 6.1.3 and 6.1.4 . . . . .	187
<b>C</b>	<b>Annex 3: surveys</b>	<b>193</b>
C.1	1.12 . . . . .	194
C.2	1.50 - 2016-06 to 2017-06 . . . . .	198
C.3	1.52 - 2017-06 to 2018-03-22 . . . . .	202
C.4	1.52.2 / 1.60 / 1.61 - 2018-03-23 onwards . . . . .	211



# List of Figures

1.1	Caillois' Ludus and Paidia . . . . .	18
1.2	Interaction cycle involving a player and a videogame . . . . .	19
1.3	A representation of the state of flow . . . . .	21
1.4	Literate world population 1800-2014 . . . . .	22
1.5	The Logo Turtle . . . . .	23
1.6	Venn diagram of video games according to Tang and Hanneghan (2007)	24
1.7	Venn diagram of video games according to Djaouti, Alvarez, and J.-P. Jessel (2011) . . . . .	24
1.8	Screenshot of <i>CellCraft</i> . . . . .	26
1.9	A Biotic Game device developed by the Riedel-Kruse lab . . . . .	27
1.10	Typical depiction of the Forgetting Curve . . . . .	33
2.1	Consequences of the Central Dogma of molecular biology: transcrip- tion and translation . . . . .	41
2.2	Registry of Standard Biological Parts: brick functions . . . . .	43
2.3	BioBricks in the transcription process . . . . .	44
2.4	The Hill function, typical gene expression from an inducible system	45
2.5	SBOL Visual: glyphs representing functional DNA sequences . . . . .	48
2.6	BioBrick sequence producing GFP. . . . .	49
2.7	Cell mechanisms modeled in Karr's model . . . . .	53
2.8	Model developed by Wortel et al. . . . .	54
2.9	Model developed by Weiße et al. . . . .	55
2.10	The Carlson Curve as of 2017 . . . . .	63
2.11	A screenshot of the MOOC IGEM High School page about BioBricks	67
2.12	Examples of YouTube videos of SB popularization . . . . .	68
4.1	Genetic devices as abilities (first gameplay level) . . . . .	84
4.2	Genetic devices as independent BioBrick sequences (second game- play level) . . . . .	84
4.3	Genetic devices as interacting BioBrick sequences (third gameplay level) . . . . .	84
4.4	Hero.Coli 1.12: screenshot of an action phase . . . . .	87
4.5	Hero.Coli 1.12: crafting interface . . . . .	87
4.6	Hero.Coli 1.12: inventory and equipment interfaces, and HUD . . . . .	88

4.7	The first two messages displayed at the beginning of the game in the 1.12 version . . . . .	89
4.8	Device tutorial in the 1.12 version . . . . .	90
4.9	RBS BioBrick tutorial in the 1.12 version . . . . .	90
4.10	Map of Hero.Coli 1.12 with highlighted checkpoints . . . . .	91
4.11	Screenshot of a <i>diff tool</i> . . . . .	96
4.12	Map of the game with highlighted path, checkpoints, and chapters .	104
4.13	Hero.Coli 1.50: screenshot of an action phase . . . . .	107
4.14	Hero.Coli 1.50: crafting interface . . . . .	108
5.1	Survey: randomized vertical position for possible answers . . . . .	113
5.2	Survey: constant horizontal position for possible answers . . . . .	113
5.3	Results of question 16 - pretest . . . . .	116
5.4	Results of question 16 - posttest . . . . .	116
5.5	Hero.Coli 1.12: pretest answers to "Which biobrick controls what is produced by the genetic device?" . . . . .	117
5.6	Hero.Coli 1.12: posttest answers to "Which biobrick controls what is produced by the genetic device?" . . . . .	117
5.7	Hero.Coli 1.12: pretest answers to "Which biobrick controls only the efficiency - level of expression - of the genetic device?" . . . . .	118
5.8	Hero.Coli 1.12: posttest answers to "Which biobrick controls only the efficiency - level of expression - of the genetic device?" . . . . .	119
6.1	Map of the game with highlighted the two blocking puzzles . . . . .	125
6.2	Pipeline of exploitation of the data from experimental subjects . . .	128
6.3	Table of correlations of demographic features and interests against scores . . . . .	130
6.4	Table of correlations of participant demographic features against their curiosity, interests, and practice . . . . .	131
6.5	Table of correlations of enjoyment against participants' characteristics	131
6.6	Table of correlations of play times against participants' self-assessed data . . . . .	132
6.7	Table of correlations of the play times against score per question and total scores . . . . .	133
6.8	Number of participants on which the correlations of figure 6.7 are based . . . . .	134
6.9	Gender of the participants kept in the study . . . . .	137
6.10	Gender of online participants (407 people, 2018-07-05 - 2018-09-19 period) . . . . .	137
6.11	Age of the participants kept in the study . . . . .	138
6.12	Age of online participants (407 people, 2018-07-05 - 2018-09-19 period)	138
6.13	Posttest answers to question 14, subquestion 6 . . . . .	139
6.14	Posttest answers to question 11 . . . . .	140
6.15	Posttest answers to question 15, subquestion 1 . . . . .	140
6.16	Learning threshold vs ratio criterion for question 15, subquestion 7	141

6.17	Learning threshold vs ratio criterion for question 12 . . . . .	141
6.18	Total posttest score vs furthest checkpoint reached . . . . .	141
6.19	Total posttest score vs furthest checkpoint reached . . . . .	143
6.20	Percentages of positive answers in pretest, posttest, and percentage increase . . . . .	143
6.21	Percentages of positive answers in pretest, posttest, and percentage increase, sorted by increase . . . . .	144
6.22	Sankey diagram of scores on BioBrick function questions using a category-lenient grading . . . . .	148
6.23	Sankey diagram of answers on the genotype-phenotype question using a strict grading . . . . .	149
6.24	Sankey diagram of answers on one induction question using a strict grading . . . . .	150
6.25	Sankey diagram of answers on three induction questions using a strict grading . . . . .	150
6.26	Change in interest in Biology . . . . .	152
6.27	Change in interest in Synthetic Biology . . . . .	152
6.28	Change in interest in Video Games . . . . .	153
6.29	Change in interest in Engineering . . . . .	153
B.1	. . . . .	186
B.2	. . . . .	186
B.3	B.1 Percentages of positive answers in pretest, posttest, and percentage increase, B.2 sorted by increase (enlarged) . . . . .	186
B.4	Matrix of correlations of demographic features and interests against scores (enlarged) . . . . .	188
B.5	Matrix of correlations of participant demographic features against their curiosity, interests, and practice (enlarged) . . . . .	189
B.6	Matrix of correlations of enjoyment against participants' characteristics (enlarged) . . . . .	190
B.7	Matrix of correlations of play times against participants' self-assessed data (enlarged) . . . . .	191
B.8	Correlation matrix of the play times against score per question and total scores (enlarged) . . . . .	192



# Synopsis

Video games (VGs), which have recently become the most ubiquitous and lucrative form of entertainment in the US (Marchand and Hennig-Thurau, 2013; Granic, Lobel, and Engels, 2014) are currently utilized and enjoyed in a variety of forms. Though VGs are often considered a symbol of non-productive activity and enjoyment, many require the user to utilize and improve upon their agility, knowledge, and intelligence to succeed. Electronic sports (eSports) players rely on their motor and decision-making skills, as well as their extensive knowledge of the game to beat their opponents. In other instances, players can spend hundreds of hours creating social structures, such as guilds (Ducheneaut et al., 2007), to form collaborative groups united by the same in-game goals - upgrading their avatars, completing a particular quest, or competing against other players. More generally, players collaborate, benchmark, hypothesize, demonstrate; they exchange tips, blueprints, strategies, and hacks, culminating in an activity called theorycrafting (Paul, 2011; Ask, 2017).

Theorycrafting is essentially conducting a thorough analysis or reverse-engineering of the mechanics and contents of a VG in order to find optimal strategies to reach an objective. An example of game with theorycrafting applications is Kerbal Space Program (KSP). In this engineering game, players explore a fictional planetary system closely resembling Earth's, using technologies closely resembling the current state of the art in rocketry. It has even achieved NASA's recognition by partnering with the agency on an in-game "asteroid redirect" mission ([https://spinoff.nasa.gov/Spinoff2015/partnership\\_1.html](https://spinoff.nasa.gov/Spinoff2015/partnership_1.html)). At first glance, KSP looks like a simple crafting and piloting VG. But some dedicated players, in a typical demonstration of theorycrafting, developed delta-v maps - maps showing the fuel cost of reaching different planets and moons - using the Tsiolkovsky rocket equation and the *vis-viva* equation, because they knew that KSP's physics simulation used Newton's law of gravitation ([https://wiki.kerbalspaceprogram.com/wiki/Cheat\\_sheet](https://wiki.kerbalspaceprogram.com/wiki/Cheat_sheet), [https://www.reddit.com/r/KerbalSpaceProgram/comments/36lu59/how\\_to\\_know\\_your\\_deltav\\_preferably\\_without\\_mods/](https://www.reddit.com/r/KerbalSpaceProgram/comments/36lu59/how_to_know_your_deltav_preferably_without_mods/)). Theorycrafting requires a rational approach to reasoning, especially the knowledge that "the same cause always produces the same effect" (David Hume, 1739-1740). It often also requires skills in math, programming, data analysis and visualization, as well as physics. To sum up, theorycrafting and the general practice of VGs can "foster scientific habits of mind" (Steinkuehler and Duncan, 2008).

VGs enable players to acquire valuable skills by playing in a safe, virtual space

and can thus be considered an example of learning through play, common in the animal kingdom (A. Y. Kolb and D. A. Kolb, 2010). Young animals exercise physically and intellectually by pretending to hunt or fight each other. Previous research equates play with learning (Singer et al., 2006), but that was only one of the reasons given by authors to explain that educators use VGs for learning. Some educators use them because playing is engaging, especially for the controversial "Digital Natives" (Prensky, 2001; Van Eck, 2006). According to some authors, Digital Natives are children born into the Internet culture, depicted as hard to engage with formal education because they are "no longer the people our educational system was designed to teach" (Prensky, 2001). Some other educators use VGs because they are a way of making use of the off-school time, more access to learning material implying more learning. Some other uses are based on the highly debated learning styles theories (Pashler et al., 2008; Willingham, Hughes, and Dobolyi, 2015), now being replaced by theories such as multimedia learning (Annetta et al., 2009), basing their success on a complementary use of sounds and visuals to support learning.

The range of new characteristics and levers the VGs provided to educators was huge. However wide this range was, the first generation of commercial off-the-shelf educational VGs of the 1980s and 1990s, the Edutainment era, failed to prove its effectiveness (Galarneau, 2005; Klopfer and Osterweil, 2013) and were caricatured as *chocolate-covered broccoli* (Amy Bruckman, 1999; Laurel, 2002). The next generations of educational VGs took the criticisms leveled against their predecessors into account by providing a stronger case for game-aided learning (Connolly et al., 2012; E. A. Boyle et al., 2016). Games started to be used in a variety of contexts, as tools with a non-leisure-based purpose. As such they are often labeled Serious Games (Susi, Johannesson, and Backlund, 2007). This marked the birth of novel types of VGs, including research and citizen science crowdsourcing VGs, Games With A Purpose (GWAPs) (Ahn, 2006), and VGs for professional training. Therapeutic games (Mader, Natkin, and Levieux, 2012) and games to raise awareness such as newsgames (Sicart, 2008) are other notable applications. Following the law of supply and demand, educational game developers targeted the most requested academic subjects, leaving niche fields deprived of applications and experimentation.

Synthetic biology (SB) was one of those niche fields that lacked VGs popularizing or teaching it until recently. SB is a recent, interdisciplinary, and applied field merging together, among others, genetics, molecular biology, and engineering, most notably genetic engineering. At its core lies the assembly of genetic sequences, employed to achieve very diverse purposes, such as minimal genome projects (Hutchison, Peterson, et al., 1999), de-extinction projects (Church and Regis, 2012), and the industrial production of chemicals using bioreactors instead of chemical plants - e.g. pharmaceutical drugs (Chris J. Paddon and Keasling, 2014) and biofuels (S. K. Lee et al., 2008). This active and promising field is also prone to security issues (Bügl et al., 2007), and consequently generates fear, be it supported by scientific evidence or not: a study supposedly proving GMO toxicity and widely

covered in the media was later retracted for lack of compliance to methodology (Séralini et al., 2014), while one of the most promising and relied upon tools of gene editing has been demonstrated to be less accurate than previously thought (Lin et al., 2014), in a context of mixed perceptions towards SB in the US (Pauwels, 2013). The general public has reacted very heterogeneously worldwide, pushing for legislation ranging from the EU's "probably strictest regulations in the world" to the laxer ones of the USA (Davison, 2010). Citizen awareness and input is needed to drive the elaboration of well-informed legal frameworks that consider the stakes, promises, and issues at play (Schmidt, Ganguli-Mitra, et al., 2009). Additionally, there are professional opportunities linked to the growth of the sector - comprising several facets like research, education, industry, arts, and leisure.

This need for SB dissemination and bi-directional communication between the general public and synthetic biologists still had to be addressed when a project to develop a synthetic-biology-themed video game, Hero.Coli, started in the CRI, in 2012. The VG format was chosen as VGs were concurrently making their debuts in research and rose to the top of leisure industries. Hero.Coli, the first VG to disseminate and popularize knowledge of SB, is the basis of the research investigation presented in this thesis, using the methodology of Design-Based Research (Wang and Hannafin, 2005; Amiel and Reeves, 2008), a methodology that relies on co-evolution of theory and practice in a continuous iterative research process.

This thesis identifies game-based learning outcomes to SB by answering the following research questions (RQs) taking into account:

RQ 1: Academic education: What are game-based learning outcomes to SB education for University students in terms of knowledge acquisition and motivation?

RQ 2: Popularization and lifelong learning: What are game-based learning outcomes to SB popularization to citizens in terms of basic comprehension and interest?

In order to ascertain accessibility, the following benefits to each of the players will be assessed:

RQ 3: Learning efficiency, motivation and player characteristics: Do players' characteristics - demographics, interests, practice - correlate to SB game-based learning efficiency in terms of knowledge acquisition and motivation?

RQ 4: Playing duration and player characteristics: Do players' characteristics - demographics, interests, practice - correlate with playing duration?

RQ 5: Player characteristics and implicit, explicit content: How do the outcomes of different pedagogical strategies compare to each other?

As multiple-choice test assessments may transform the experience into *chocolate-covered broccoli* by breaking the flow of the game (Valerie J Shute, 2011), they should be avoided when possible. When learning does not break the

flow of the VG, it becomes *stealth learning* (Paras and Bizzocchi, 2005), an objective that was set when designing Hero.Coli. Therefore we set out to answer the following questions:

- RQ 6: Quiz-based assessment and automated tracking: How comparable are learning metrics computed from questionnaires and from automated remote tracking data? Can quiz-based assessment be replaced by automated tracking?
- RQ 7: Threshold effect: Is there a threshold effect in the game, *i.e.* a point in the game after which no significant additional outcome is measured?

Herein, we demonstrate the elaboration of Hero.Coli to teach and popularize SB based on research in the field of educational VGs, and evaluate it with regard to the aforementioned 7 questions.

To support the purpose of this thesis, part I is an overview of the context in both educational VGs and SB fields. First, by showcasing recent uses and discoveries of VGs for science and education. In particular, the characteristics of educational and awareness VGs are delineated. Different techniques are listed, along with their advantages and drawbacks, to prepare and justify design choices depicted later in this thesis. Part I also broadly presents the field of SB and its specificities. Evaluating a SB-teaching tool indeed implies to first acknowledge SB's approach to living systems, its principles, and its goals. Finally, concluding this part, pedagogical objectives and assessment metrics are set.

Part II focuses on the experimental setup, providing details about Hero.Coli. Its origins and creation process prior to this study are outlined, followed by its re-purposing, re-design, enhancements, and later additions - such as the analytics system. Choices concerning realism, story, game level design are discussed regarding pedagogical objectives and research results listed in the previous part.

Part III describes the data gathering campaigns based on surveys and remote tracking among students and citizens, and the analyses which were implemented in order to answer the research questions. Different iterations were necessary to achieve the goals that were set, narrowing down on the intended scope of the study. The first surveys were more game-oriented, to address basic acceptability and usability issues. Subsequent surveys highlighted misunderstandings and educational successes. Finally, precise learning outcomes were assessed.

Part IV draws conclusions from those data, confirming the potential for VGs in the teaching and popularization of SB, with listed limitations and suggested best practices. New applications and further research axes are also proposed, building upon this study to suggest new pedagogical objectives and appropriate techniques to teach them. In particular, game elements that were not explicitly demonstrated and explained, or content that was not paid enough attention to, could be selectively introduced to the user, while simplifications made on the simulator could be lifted to enable access to higher-level SB knowledge and applications.

**Part I**  
**Introduction**



# Chapter 1

## Video games and Learning

This chapter presents the state of the art in the literature in video games (VGs) used for learning, also called digital game-based learning. The field is presented in relation with neighboring notions in digital learning to pinpoint its characteristics, objectives, and means of action. In this thesis, we will explore the lead of using video game-based learning as a means to ease up the task of teachers in formal education, and also as a tool for lifelong learning and popularization. Education is indeed not restricted to formal education anymore: students are encouraged to be autonomous. The Twentieth Century Skills (Dede, 2009), a list of skills deemed crucial in today's and tomorrow's world by organizations such as the OECD, comprise autonomy as a core value. Additionally, popularization through VGs is a new means to raise awareness among a large audience about topics from geopolitics to societal challenges (Jacobs, Jansz, and Hera CondePumpido, 2017).

### 1.1 Generalities about video games

#### 1.1.1 Game

Authors have produced a plethora of definitions for games, some of them contradicting each other. The central role of rules to delineate a game are central and common to most definitions. One of the most seminal definitions is Roger Caillois' restrictive definition of a game (Caillois, 1958) as an activity that is voluntary, confined in time and space, uncertain *i.e.* driven by player decision or chance (randomness), unproductive, regulated, and fictitious, *i.e.* in which the willing suspension of disbelief is necessary. Caillois also made the distinction between *ludus* (free play) and *paidia* (constrained game) (Caillois and Halperin, 1955) (see figure 1.1).

Huizinga proposed the following definition:

“Summing up the formal characteristics of play we might call it a free activity standing quite consciously outside ‘ordinary’ life as being ‘not serious,’ but at the same time absorbing the player intensely and utterly. It is an activity connected with no material interest, and no profit

	AGÔN (Competition)	ALEA (Chance)	MIMICRY (Simulation)	ILINX (Vertigo)
PAIDIA Tumult Agitation Immoderate laughter	Racing Wrestling Etc. } not regulated Athletics	Counting-out rhymes Heads or tails	Children's initiations Games of illusion Tag, Arms Masks, Disguises	Children "whirling" Horseback riding Swinging Waltzing
Kite-flying Solitaire Patience Crossword puzzles	Boxing, Billiards Fencing, Checkers Football, Chess	Betting Roulette		Volador Traveling carnivals Skiing Mountain climbing Tightrope walking
LUDUS	Contests, Sports in general	Simple, complex, and continuing lotteries*	Theater Spectacles in general	

Figure 1.1: Caillois' Ludus and Paidia

Source: Adapted from Caillois and Halperin (1955), <https://dredtabletop.wordpress.com/2016/02/07/caillois-and-theorytypes-of-game-players/>

can be gained by it. It proceeds within its own proper boundaries of time and space according to fixed rules and in an orderly manner. It promotes the formation of social groupings which tend to surround themselves with secrecy and to stress their difference from the common world by disguise or other means." Huizinga (1938)

It adds to the futility of games its frivolousness, while introducing an immersion aspect, prefiguring the concepts of engagement and flow. These definitions have however been challenged by the new uses of games: games are not unproductive anymore as virtual items and characters of Massively Multiplayer Online Role Playing Games (MMORPGs) can be exchanged against real currency, creating a market of fluctuating virtual goods. Games can also be productive when they teach valuable skills. These definitions also do not apply for sandbox games. In these games, there are no winning objectives and the set of rules are usually restricted to the laws of physics, and a few other rules preventing the game from reaching a stuck or uninteresting state - or from crashing. For instance, broken elements are usually removed to prevent cluttering. A machine stuck upside-down is usually automatically set back to its functioning position. But there are usually no *points* nor *achievements* or *victory positions* to be taken in sandbox games. One could argue that the problem only stem from an improper use of the term "game" (*paidia*, constrained game), instead of "toy" (*ludus*, free play). However, sandbox games such as the free-roaming MMORPG *Eve Online* (Carter and Gibbs, 2013) make the boundary between *ludus* and *paidia* even blurrier. *Eve Online* is a successful multiplayer game in which it is impossible to win or lose,

and in which, contrary to the MMORPG World of Warcraft, there is no storyline to follow. Players are free to pursue the goals they want. Other genres of games challenge Caillois' definition, such as contemplative games exemplified by the game *Dear Esther* (Morisset, 2014) which is a story narrated as the player explores an environment. These challenges have led authors to search for new definitions of games in the recent years (Juul, 2003).

All of the examples given here are VGs but these definitions also apply to other genres of games such as board games, role-playing games, and escape games. VGs distinguish themselves only by the hardware used (see figure 1.2). In role-playing games, the constraints, rules, and game mechanics are managed by a game master who uses dice and rule books; in board games, they are managed by an optional game master and by items such as cards and pieces; in escape games the room itself manages most of the interactions while a game master monitors the game backstage. The inputs are either manual or through an interaction with the game master, and the displaying system uses real items, such as books, cards, or specific props. Of course, some escape games use computers but their use is intradiegetic: these computers are part of the universe of the game itself and not a way to display or simulate part of the game.

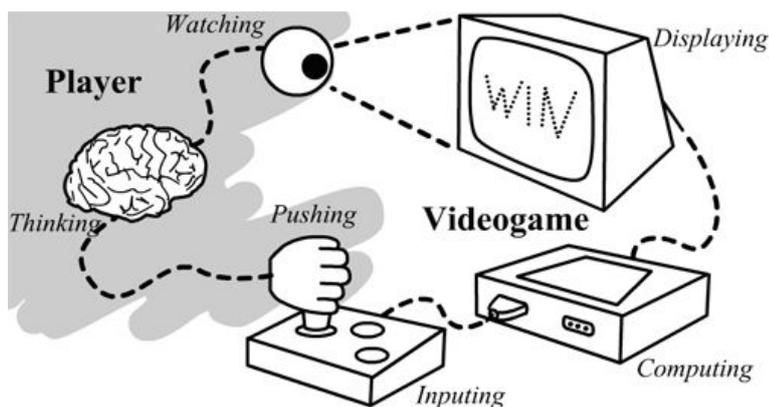


Figure 1.2: Interaction cycle involving a player and a videogame  
 Source: *Djaouti, Alvarez, J.-p. Jessel, et al. (2008)*

### 1.1.2 Gameplay and Gameplay Loops

Compared to the interface, a well-defined and ubiquitous element in digital platforms, "Gameplay" is a more nebulous term" (Juul and Norton, 2009). Jesper Juul defines gameplay as:

[...] not how a game looks, but how it plays: how the player interacts with its rules and experiences the totality of challenges and choices that the game offers. In a technical sense, gameplay always concerns the player's interaction with the underlying state of a game, and gameplay

is typically used to describe the specific experience of interacting with the game, independently of graphics, fiction, and audio, even if the total player experience is influenced by these other design elements. Ryan, Emerson, and Robertson (2014)

The key elements here are interaction and game state: gameplay is how the player controls the state of the game.

Gameplay loops are built upon the notion of gameplay and are also defined differently across different contexts (Perron and Arsenault, 2008; Guardiola, 2016). Definitions stemming from game studies often focus on the atomic loop of interaction and display (see figure 1.2). Definitions stemming from the video game industry include the Objective-Challenge-Reward loops of different time scales usually called micro and macro gameplay loops. A typical example used to illustrate these loops is *Super Mario*: the micro gameplay loop involves advancing to the right while avoiding dangers and collecting bonuses, on a time scale of a few seconds. The macro game loop involves completing the whole game by unlocking all the levels one by one, on the time scale of a whole game. Intermediary game loops are often also defined, here involving the completion of a level, on a time scale of a few minutes. In the *Sim City* series, a series of games where the player builds and manages a city, the micro gameplay loop involves building, and monitoring emergencies; the intermediary gameplay loop consists in managing the yearly budget by balancing the expenses and planning for the next year; and the macro gameplay loop consists in reaching the objectives set in the scenario: reaching a population level, managing the city until a given date, or gathering a given amount of money.

Gameplay and gameplay loops are useful tools to describe and study games alongside the notion of game genres. Depending on the way game genres are defined, they can be listed from 4 - action, adventure, strategy, puzzle (Rollings and E. Adams, 2003) - to forty-two (Wolf, 2001). Actually, the number of genres is ever growing as new game genres may be created every year if they are based on new technologies: new displays (Virtual Reality systems), new controllers. This unreliable definition of game genres has been questioned by some authors (Apperley, 2006).

Gameplay, gameplay loops, and game genre are also at the core of the role of the game designer, alongside the narration, the scenario, the story of the game, its characters and interactions. Aesthetics and graphics, however central in the identity of a game, are not among the responsibilities of the game designer. Game design also often comprises level design - the way game elements such as traps, bonuses, platforms, and doors are arranged in a level. By creating and tuning all of these elements, the game designer produces an experience that will motivate or engage the player. In the literature, motivation and engagement relative to a game are often used interchangeably to describe a state of emotional involvement and investment in a game, with an expected emotional outcome. Engagement is "the willingness to have emotions, affect and thoughts directed towards and determined by the mediated activity" (Patrice Bouvier, Lavoué, et al., 2013). The same authors expand this definition of engagement in another publication:

We consider the engagement of a player as the desire to have emotions, affect and thoughts directed to and determined by the mediated activity. This "engaged" state means in particular that:

- The game arouses emotions (such as joy, pride, accomplishment, enjoyment or frustration) for the player.
- The game occupies the thoughts of the player during the gaming sessions but also outside.
- The player wishes to continue playing.

Thus, the engagement requires an intellectual and emotional investment from the player which goes beyond the discovery phase of the game. Patrice Bouvier, Sehaba, and Lavoué (2014)

Motivation is sometimes described as more related to a state of mind characterized by a willingness to participate to an activity while engagement is described as more of a behavior.

Engagement is also presented in the theory of *flow* (Csikszentmihalyi, 1997; Jenova Chen, 2011), the need to maintain a constantly balanced challenge to never break the momentum and immersion (figure 1.3). Indeed, games that are too hard frustrate their users while games that are too easy bore them. That is why levels of difficulty are so widespread among VGs: to manually adjust the challenge.

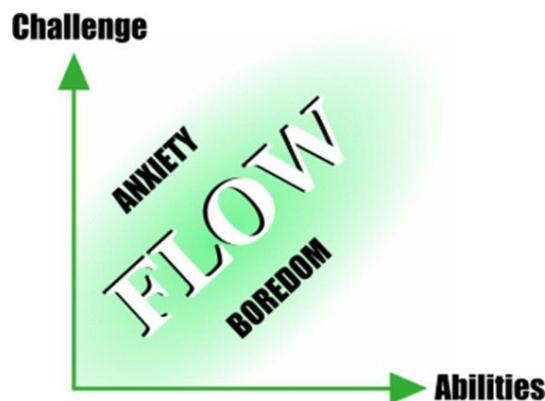


Figure 1.3: A representation of the state of flow

Source: <http://jenovachen.info>

Having presented games, VGs and notions necessary to describe and study them, we will focus on digital game-based learning, *i.e.* learning based on VGs.

## 1.2 Digital game-based learning

### 1.2.1 The need for game-based learning

Education has become a key state priority since the industrialization period all over the world (Carl, 2009), and is therefore subject to close scrutiny and optimization. Great efforts have been deployed to enforce its democratization: literacy has risen from 12% to 85% in the world from 1800 to 2014 (see figure 1.4).

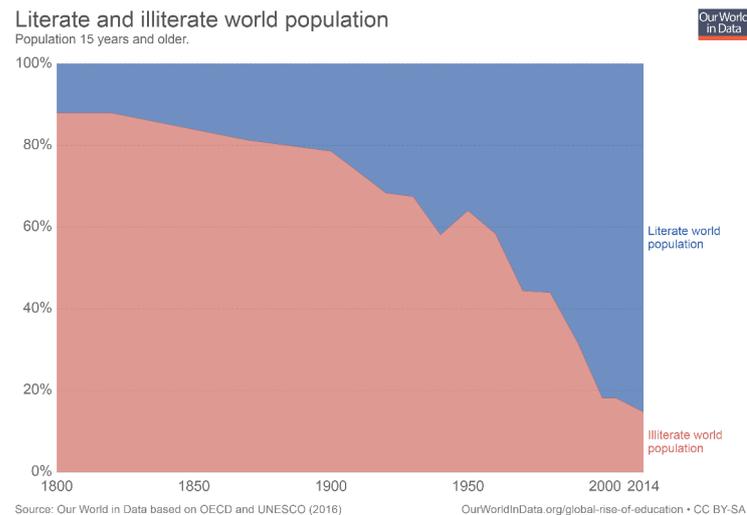


Figure 1.4: Literate world population 1800-2014

Source: Data calculated by Max Roser for Our World In Data, 2016, from OECD and UNESCO data <https://ourworldindata.org/literacy>

The need for affordable large-scale education has driven new pedagogies and new technologies to be developed alongside. New pedagogies stress that the interaction between teacher and students is paramount, and that pupils have to take part actively in the learning process. These ideas are further developed and exemplified in the *constructivism* approach initiated by Piaget (Piaget, 1970), which focuses on the process of creation of knowledge by humans through the interactions between new experiences and prior knowledge. Other approaches rely on the use of new technologies. For instance, Edison promoted his invention, the phonograph, by describing how its recording and playing features enabled students to replay the courses at will (Symes, 2004). Other applications of technological innovations to education include cinema, radio, TV, VGs, online courses, smartphone-based educational applications. Chronologically, VGs are therefore one iteration in a long series. None of them has yet proven efficient enough to curb the cost and demonstratively ease up the process of education at school - *formal education* (Mayes, n.d.), and education is still considered by many authors to be in turmoil. Some justify it by the attitude of the controversial "Digital Natives"

(Prensky, 2001; Van Eck, 2006), children born into the Internet culture, depicted as hard to engage with formal education because they are "no longer the people our educational system was designed to teach" (Prensky, 2001). Some teachers justify it by the pressure governments apply to save costs with budget cuts, resulting in insufficient time and personnel dedicated to teaching among other reasons (Ingersoll, 2003; Sutchter, Darling-Hammond, and Carver-Thomas, 2016).

One of the first uses of *gamified programs* for learning dates back to the sixties, with the work of Wally Feurzeig and Seymour Papert on teaching programming to kids through a virtual turtle using the Logo high-level programming language (Feurzeig et al., 1969; Papert, 1980) (figure 1.5).

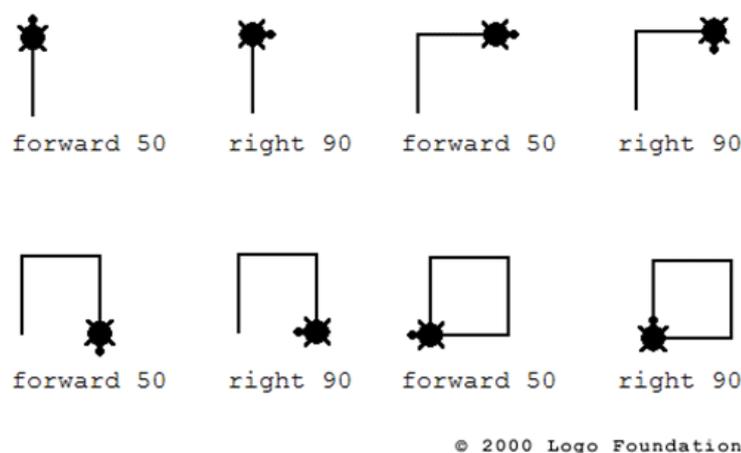


Figure 1.5: The Logo Turtle  
 Source: Papert (1980)

Another example of early implementation is the PLATO (Programmed Logic for Automated Teaching Operations) program (Bitzer, Braunfeld, and Lichtenberger, 1961) which ran from 1960 to 1985, at the University of Illinois. These very first programs to be used to teach to kids were *gamified* programs, *i.e.* programs that feature characteristics usually considered to be idiosyncratic of games in order to increase engagement. Reward systems - points or unlocked functionalities and content - are examples of those game-like characteristics. A visual identity, visual elements, or graphic assets set in a range of codified fantasy universes are also regarded as pertaining to games. Cartoon characters, space, pirates are among the most frequent game universe tropes.

## 1.2.2 Applications of digital game-based learning

Different VGs target different audiences, prioritize differently realism or fun. That is why it is important to classify VGs according to their context of creation and use. Figure 1.6 from Tang and Hanneghan (2007) is a Venn diagram representing the use of VGs and closely related programs in Education. There is no consensus

on the terminology but this one is among the most widespread internationally - for example in France by Djaouti, Alvarez, and J.-P. Jessel (2011) depicted in figure 1.7, which adds notions such as "Serious Gaming" and "purpose-shifted VGs".

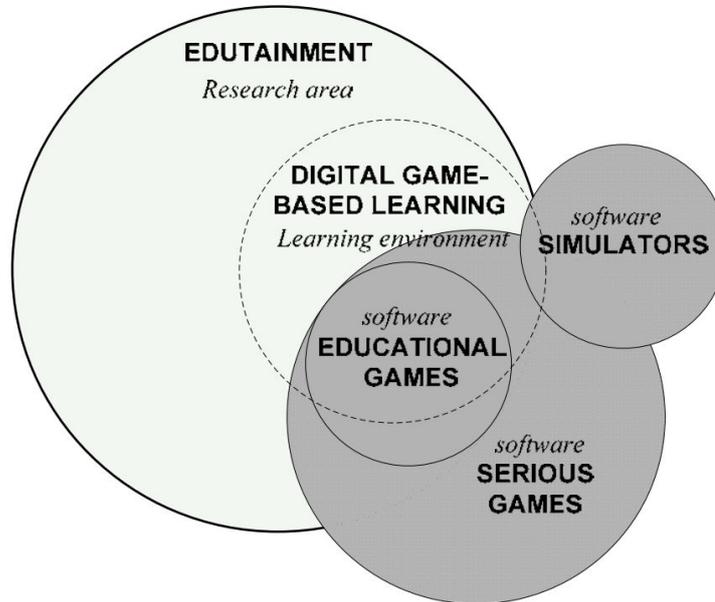


Figure 1.6: Venn diagram of video games according to Tang and Hanneghan (2007)

Source: Tang and Hanneghan (2007)

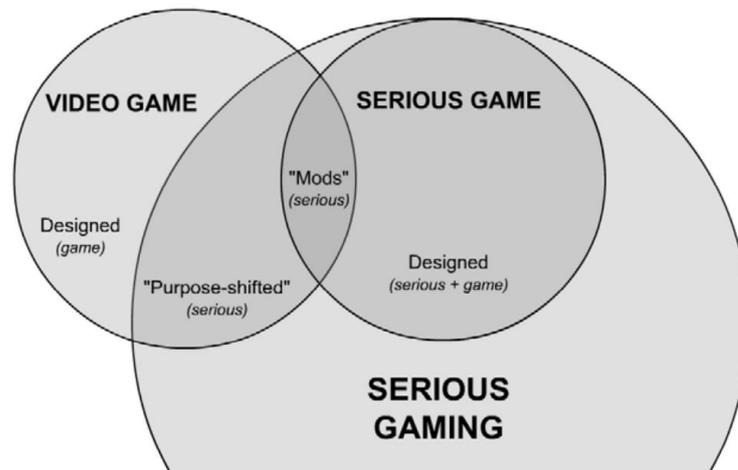


Figure 1.7: Venn diagram of video games according to Djaouti, Alvarez, and J.-P. Jessel (2011)

Source: Djaouti, Alvarez, and J.-P. Jessel (2011)

The terminology in game-based learning is therefore:

- Digital Game-Based Learning: the use of VGs in Education.
- Simulators: simulators in their wider sense and use.
- Educational Games: games intending on teaching first, particularly used in an academic setting. The whole genre is sometimes labeled "Edutainment".
- Serious Games: games developed for and used in professional training; often also conflated with games with a scientific intent, Citizen Science games (crowdsourced and gamified research projects), therapeutic games, and so on in a category often called Games With A Purpose (GWAPs);
- Serious Gaming: using VGs to train or educate, be the VGs serious games, educational games, or Commercial Off-The-Shelf digital games (COTS) games. In the last case, the games have had "Mods"<sup>1</sup> developed to this end or have been "purpose-shifted" (Djaouti, Alvarez, and J.-P. Jessel, 2011).

In this thesis, I will adopt this terminology, with subcategories. I will add the popularization games - games made by specialists and professionals to raise awareness about their field - and thematic games - COTS games on a scientific or professional topic with a imbalance in favor of fun against realism and accuracy. Serious Games will be split in educational games (for academic or professional purposes), therapeutic games, Citizen Science games and other research games.

### Educational games

Educational games are usually developed or funded by academic institutions, and usually struggle to succeed as COTS. The aim is to give a new way of learning to students, to offer novelty. The aim is also sometimes to give an *additional* way of learning to students, according to the disputed theory of *learning styles* (Pashler et al., 2008; Willingham, Hughes, and Dobolyi, 2015), now complemented by theories such as multimedia learning (Annetta et al., 2009).

**In biology** Cellcraft<sup>2</sup> is an educational game about cell biology (Dunlap and Pecore, 2009). It integrates higher-education-level content - cf figure 1.8 - into convincing gameplay, making it a perfect example of educational game using intrinsic motivation. Gaming sequences are not mere rewards of learning, learning is blended into gaming.

---

<sup>1</sup> Mods are extensions of VGs often created by players themselves to add functionalities or environments to an existing game.

<sup>2</sup> On the Internet: <http://cellcraftgame.com/>, <https://www.kongregate.com/games/CellCraft/cellcraft>. The title of the game is a direct reference to the famous games Warcraft and Starcraft by Blizzard.



Figure 1.8: Screenshot of *CellCraft*

Source: Science game center <http://www.sciencegamecenter.org/screenshots/57>

Other examples include a simulator of the lactose operon (Esmaeili et al., 2015) and Synmod (Schmidt, Radchuk, and Meinhart, 2014), a game to train and memorize the amino-acids and their properties.

Other projects include tangible elements, even going as far as including SB kits in the *Bixel* project<sup>3</sup>.

### Popularization games

They are the equivalent of *Jurassic Park*: an entry point for paleontology but with a few inaccuracies. They intend on teaching a few notions of the discipline being presented: definitions, mechanisms, methods, problems, perspectives - the same way newsgames try and raise awareness about issues such as the Syrian Civil War and the War in Darfur<sup>4</sup>, without caricature or simplification. The realism/fun balance is clearly in favor of fun, but realism is still important.

Among the most iconic popularization games are history games of the Age of Empires (AoE) and Civilization series, and the space exploration game Kerbal Space Program (KSP) introduced in the synopsis. They all have their inaccuracies and approximations - unrealistic army sizes, time scales, physical scale, logistics, simplified physics, ... - but on the topic they intend to popularize they maintain a high level of faithfulness. AoE games feature internal encyclopedias to

<sup>3</sup> <http://www.imperial.ac.uk/news/183377/bio-computer-powered-jellyfish-dna-plays-tetris/>

<sup>4</sup> Among the most successful newsgames are *Endgame: Syria*, Auroch Digital Ltd, 2012, and *Darfur is Dying*, interFUEL, LLC, 2006.

learn more about the history of civilizations playable in the game. Civilization games feature an impressive amount of game elements inspired from reality, from civilization-specific army units to national landmarks, historical events, technological breakthroughs, political power structures and ideologies. KSP is a crafting game comprising hundreds of rocket parts using real and experimental technologies. KSP also boasts collaborating with NASA to provide scenarios inspired from real NASA programs, such as the Asteroid Redirect Mission.

**In biology** Some research teams have also explored another way to get the general public interested in biology. They have developed VGs using equipment used in experimental microbiology, called *biotic games* (H. Riedel-Kruse et al., 2011; Harvey et al., 2014). They allow players to interact with real living microscopic organisms. The goal is to have non-scientists discover life at the microscopic level, experiment with the tools and mechanisms at hand - lenses, growth medium, systems to get the cells to move -, find out facts about these microorganisms by themselves, develop their curiosity, and start discussing with scientists. For instance, the Riedel-Kruse lab, a bioengineering lab in Stanford, California<sup>5</sup> has developed complete physical devices (cf figure 1.9) and code making it possible to control phototactic<sup>6</sup> unicellular organisms and integrate them in games (Cira et al., 2015; Washington et al., 2018).

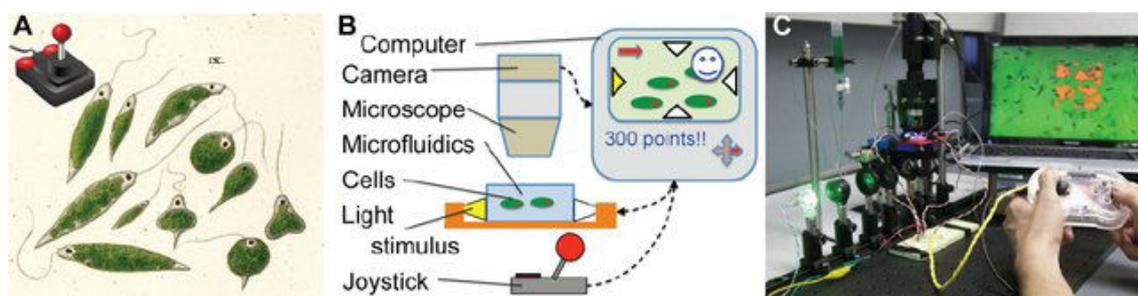


Figure 1.9: A Biotic Game device developed by the Riedel-Kruse lab  
 Source: Cira et al. (2015)

In Europe, researchers Roland Van Dierendonck<sup>7</sup> and Wim van Eck<sup>8</sup> developed similar devices.

Other games aiming at popularizing practices in biology through the mixed use of digital games and living organisms are Bactman Adventures (<http://2015>).

<sup>5</sup> <https://web.stanford.edu/group/riedel-kruse/index.html>.

<sup>6</sup> Phototactic organisms are attracted to or repelled by light.

<sup>7</sup> Roland Van Dierendonck leads the BioHack Academy in Waag Society, an institute for art, science and technology in Amsterdam, Netherlands. Personal page: <https://waag.org/en/roland-van-dierendonck>

<sup>8</sup> Wim van Eck is PhD candidate at Leiden University. Personal page: <https://www.wimeck.com/about/>

igem.org/Team:IONIS\_Paris/Description) and Colisweeper ([http://2013.igem.org/Team:ETH\\_Zurich](http://2013.igem.org/Team:ETH_Zurich)), both developed by synthetic biologists participating in iGEM, a synthetic biology competition presented in chapter 2.

### Thematic games

They are the equivalent of *Star Wars*: an entry point for aerospace engineering or astrophysics, but with a lot of inaccuracies. Spore is a COTS game that allows the player to lead a species from the microscopic scale to an intergalactic empire in 5 chapters. Spore was advertised as a game that would teach evolution in a context of growing influence from creationists. In the first two chapters, the Cell stage and the Creature stage, the character that the player controls evolves and acquires traits that makes it able to beat other competing creatures. The Cell stage takes place in an aquatic environment while the Creature stage takes places on land. Later on, the player controls a tribe comprising several individuals of this species and leads it through various cultural and technological steps. The issue is that, first, life appears by panspermia in the game. It is not an issue scientifically, but pedagogically, this eludes the question of how life appeared. Then, in the Cell stage, there is neither random mutation, nor reproduction. The player controls one creature that they *designed* using parts, like fins, eyes, and fangs that they assembled. There is only a kind of "selection" through gameplay: the ill-designed creatures will be killed off by predators, however dexterous the player is. This assembly of body parts goes on in the second stage. Spore has sold a million copies, but has also been criticized for presenting an intelligent design vision of evolution although evolutionary biologists were interviewed during the elaboration of the game. A group of scientists assessed the game's accuracy: "The game flunked evolutionary biology outright with an F. According to Gregory and Eldredge, "Spore has very little to do with real biology"<sup>9</sup> (Bohannon, 2008; Owens, 2012). The game is still usable as Serious Gaming tool though (Schrader, Deniz, and Keilty, 2016). Teachers can focus on the positive aspects and themes - fitness in an environment, predation, survival of a species - and train the students to reflect critically on less accurate aspects of the game.

### 1.2.3 Other Serious Games and GWAPs

This thesis does not focus on other applications of serious games and GWAPs. However, the interaction between research and VGs should be underlined by presenting brief additional examples.

---

<sup>9</sup> Complete report on the *Wayback Machine* website: [/web/20081028184213/http://scienceguild.org/wiki/index.php?title=Spore](http://web/20081028184213/http://scienceguild.org/wiki/index.php?title=Spore)

## Citizen Science

Citizen science games are crowdsourced games to do research<sup>10</sup>.

Players either make scientific progress themselves by producing, classifying, deciphering, or analyzing data, or give computing time on their digital devices for remotely-controlled programs to automatically process data. See table 1.1 for a list of Citizen Science projects.

Game name	Field	Data format	Task
GalaxyZoo	astronomy	deep-space pictures	classification
FoldIt	biochemistry	amino-acid sequences	protein folding
Eyewire	neurology	pictures of brain slices	connectome mapping
Fraxinus, Phylo	genetics	DNA sequences	sequence alignment
Volunteer Science	social sciences	theoretical and practical problems, thought experiments	filling in surveys

Table 1.1: Citizen Science games

Source: <http://www.citizensciencecenter.com/citizen-science-games-ultimate-list/>, <https://citizensciencegames.com/>, and the websites of those projects

In the scope of this thesis, the most interesting and inspiring project is FoldIt (Cooper et al., 2010). This iconic game has its players try and fold proteins to reach their most stable spatial conformation - see chapter 2 section 2.1.1 for a presentation of the protein production process. In short, proteins are strings of elements called amino-acids. Protein sequencing - determining the sequence of amino-acids comprising a protein - is easy, but protein folding - determining the 3d conformation of a protein - is hard. As protein conformation is related to the function of a protein, biologists had listed numerous protein sequences but did not know what their function was when the FoldIt project started. The fact that crowdsourcing was required to only start tackling protein folding comes from the fact that the protein folding problem is famous for its combinatorial size (Levinthal, 1969). Levinthal has estimated that a 150-amino-acid long protein can fold in  $10^{300}$  configurations, but takes only microseconds to reach equilibrium<sup>11</sup>. This astronomical number of configurations is moreover only for *one* 150-amino-acid long protein, among all the existing proteins, although they comprise on

<sup>10</sup> For a list of citizen science games: <http://www.citizensciencecenter.com/citizen-science-games-ultimate-list/> and <https://citizensciencegames.com/>

<sup>11</sup> Digital version of Levinthal's talk and paper: <https://archive.is/EM2n>.

average 300 amino-acids for *E. coli*, with up to 2367 amino-acids<sup>12</sup>. A computer program cannot *brute-force* a problem with  $10^{300}$  configurations. On the other hand, humans are better at discovering patterns without any prior knowledge on the subject - at least as of 2018: recent progresses in AI has lead programs to discover patterns by themselves, with methods such as unsupervised learning. Back when FoldIt started, at any rate, crowdsourcing was the best option, because, for instance, humans could identify regions of proteins that tend to fold into helices, or pairs of regions which tend to bind to one another. Later, these patterns were fed as heuristics to protein folding programs, making it possible to accelerate the process and avoid an ineffective brute-forcing.

FoldIt has inspired many projects, such as EteRNA (J. Lee et al., 2014), Nanocrafter (Barone et al., 2015) and Phylo (Kawrykow et al., 2012), which deal with RNA folding, DNA folding, and DNA sequence alignment and phylogeny.

It is important to note that playing these games always induce some amount of learning. FoldIt players know about proteins and the protein-folding problem. But those VGs are not tuned to optimize learning. They are tuned to optimize usability: in the interest of crowdsourcing players have to be experts of the game rather than of the topic. In the case of FoldIt, players are experts of folding tools, camera manipulation, communication on the forums, pattern detection.

Crowdsourcing has also been implemented in COTS games: Vacnet on CS:GO and slate analysis in Eve Online.

### Therapeutic games

Therapeutic games are also called Health games. VGs have been used successfully to treat mental health issues (Wilkinson, Ang, and Goh, 2008). Monitored by their doctor, patients confront the object of their phobia, or PTSD-triggering situations, in a harmless, controlled virtual environment. Another medical use of VGs is rehabilitation therapies, through *exergaming*. Using specifically-designed game controllers or consumer solutions such as the Wii controller, these exergames act as a guide for patient physical exercise by having the patient do certain movements on a certain rhythm. A tennis game can be prescribed for shoulder rehabilitation for instance. Bamparopoulos et al. (2016) also added data tracking and Citizen Science to crowdsource measurements performed during rehabilitation. Diagnosis games are also often included in this category. Those are VGs aiming at gathering data that a doctor may take into account into their diagnosis, or games integrating a diagnosing algorithm.

#### 1.2.4 Closely-related genres

Other types of solutions for digital learning have been developed that share characteristics with digital game-based learning: gamified learning apps and simulators.

---

<sup>12</sup> Source: Harvard University's online database Bionumbers <http://bionumbers.hms.harvard.edu/bionumber.aspx?s=n&v=3&id=108985>.

### Gamified learning programs and apps

In the 1960s, Wally Feurzeig and Seymour Papert conducted their first experiments on teaching programming to kids. Kids had to program the behavior of a virtual turtle using the Logo high-level programming language (Feurzeig et al., 1969). The instruction given by the teacher was to draw a given pattern using the on-screen turtle, on which was attached a virtual pen. This was inspired by *turtle robots* of the 1940s used for the instruction and training of engineers<sup>13</sup>. Note that this experiment already used gamification, as stated in section 1.2.1: *gamified* programs are programs that feature characteristics usually considered to be idiosyncratic of games in order to increase engagement. In the case of Feurzeig and Papert's experiment, the *turtle* and the *drawing* elements of the pedagogical scenario could easily be changed for anything else, and they belong to the mental universe of leisure for children (a turtle is a cute innocent animal, drawing is a typical kid activity). For instance, the scenario could have been: "as a carpenter, you need to saw wood planks in the designated shapes in order to build new furniture using a programmable machine", or "as a cloth maker, you need to sew along these given patterns using a programmable machine". An additional advantage of using a turtle was the possibility for children to imagine themselves being the turtle, making orientation, rotation, movement easier. Instead of having kids confused about absolute or relative coordinates of an abstract cursor, this implementation made it possible to embody the turtle and simplifying the problem into "turning ten degrees to the right, then walking two steps". This experiment yielded great educational successes, but raised the question of translating educational successes to concrete effects in practical uses: children learned how to move a turtle, but not necessarily how to use coordinate systems and how to code in general. Some notions were *introduced* - procedural programming, code, geometric angles, and some aspects of rationality - but not necessarily *mastered*. The *transfer* of knowledge may not be as spectacular as expected: children became good at drawing with a programmable turtle but that is all.

Another example of early implementation is the PLATO (Programmed Logic for Automated Teaching Operations) program (Bitzer, Braunfeld, and Lichtenberger, 1961) which ran from 1960 to 1985, at the University of Illinois. Both these examples yielded encouraging results but could not scale due to the then scarcity of computers. When personal computers and laptops became popular, Seymour Papert lead several initiatives to have every child, every student get a computer (Stager, 2016).

Compared to gamified learning apps, VGs benefit from advantages of computer-based systems while adding game advantages. They build on scalability, connection to the Internet, multimedia by adding gamification and game aspects. The Internet made it possible to gather user data and to patch games remotely and efficiently. It enabled a quick feedback loop between users and developers, based on different techniques such as A/B testing, that lead to widely adopted conventions

---

<sup>13</sup> Source: [http://snuet.com/CML/C03/C03\\_02.html](http://snuet.com/CML/C03/C03_02.html)

in GUIs and mechanics in apps and games. They have been proved to enhance motor skills in terms of precision and speed, spatial orientation, 3d representation, engagement. As an educational tool, the most interesting characteristics of VGs are engagement, as engagement is correlated with learning.

However, gamified learning apps have also found their audience. From the 2010 onwards, smartphones with computing power and graphics reached desktop computer levels of the previous decade, achieved a wide commercial success and became widespread. The trend of *apps* - smartphone applications - took on and allowed for diversified implementations, from tools inherited from the digital personal manager era, to the leisure apps - games, music, audiobooks, social networks... In addition to leisure time, these apps could be used in any time slot that was until then boredom, lost time, *unproductive* time such as waiting lines and public transportation. Developers of gamified learning apps used as marketing argument that this lost time could be made productive by spending it learning. Indeed, practicing everyday a little is better than practicing once intensively. Moreover, some apps also use the principle of *spaced repetition* by scheduling reviews: they schedule when the learner should review learned content, but also the precise content of the review. Each learned item - for instance, a word in a foreign language and its translation - has its own timer obeying the principle of spaced repetition (Kang, 2016; Ferguson et al., 2017) : reviews have to happen at increasingly long periods. This is what Ebbinghaus controversially discovered in 1885 (Ebbinghaus, 1885) - he was his only experimental subject. More recent studies have confirmed the overall shape of the curve now known as the Forgetting Curve: learning outcomes can be reinforced through frequent review (Bailey, 1989; Averell and Heathcote, 2011). A typical representation of the curve is shown in figure 1.10. The blue curve represents a measure of the percentage of knowledge retained from the learning event at  $t_0 = 0$  which can be assessed at any given point in time. The relearning or reviewing events at  $t_1 = 1$  day,  $t_2 = 3$  days,  $t_3 = 10$  days,  $t_4 = 30$  days show that ulterior reviews not only put the assessed level of memorized content back to the maximum, but also reduce the rate of forgetting over time. This reduced rate means that those reviews can be more and more spaced out.

There are many actors in the gamified learning app market: DuoLingo, Memrise, Tinycards, SoloLearn, Clozemasters... They usually focus on language learning, but can actually be employed to learn geography, science, coding, or even yoga positions. These apps are characterized by the fact that they state that they are serious apps (not games), while using game features such as points, achievements, avatars. MOOC providers and platforms also use apps to broadcast their content, such as Khan Academy and Coursera.

### Simulations and simulators

Simulators used for learning and training rely on the accurate reconstitution of real-life events and situations. Exposing learners to these situations in a safe, consequence-less, cheaper setting allows for reduced stress and helps getting accustomed. Simulators in their broader definition may only include recreated premises

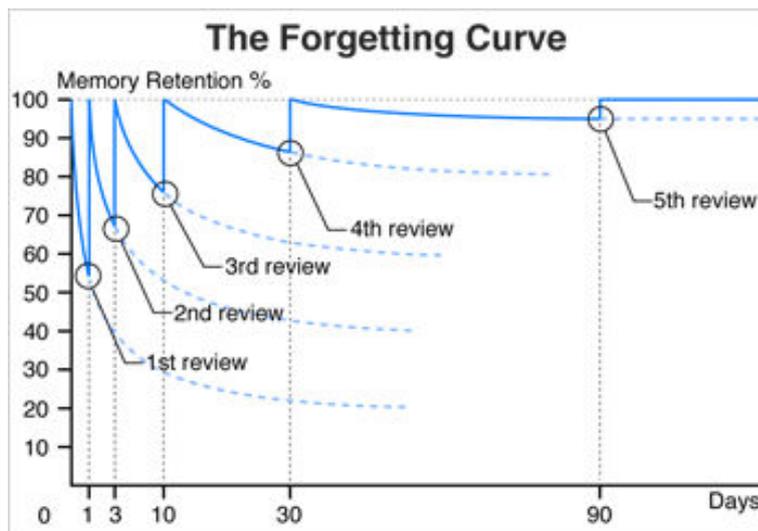


Figure 1.10: Typical depiction of the Forgetting Curve

Source: <http://www.criticaltosuccess.com/how-to-build-super-memory-business/>

coupled with role-play, in order to practice procedures, such as emergency protocols like evacuations or the treatment of the wounded. But usually simulators now include computer simulations which will handle the role-play rules, the physics simulation, and the visual and audio rendering. Some systems are hybrid. Professional flight simulators feature a faithfully recreated cockpit, with windscreens replaced by digital displays, showing a real-time realistic simulated flight environment. The digital simulation behind this system is extremely faithful to reality. It enables pilots to safely repeat routine and emergency procedures, while also enabling air crash investigators to recreate the exact conditions of a crash. Simulators have been tested and proven as efficient teaching systems in academic and professional settings (Freitas and Maharg, 2011). For instance, scientific popularization websites such as MinuteLabs.io<sup>14</sup>, closely linked to the notorious YouTube channels of science popularization MinutePhysics<sup>15</sup> and MinuteEarth<sup>16</sup> enable web visitors to run physics-based simulations. Interestingly, they have also been adopted by the leisure industry. The trends of *sandbox games* and *crafting games* in the video game industry demonstrate the attractiveness of simulators to video game users.

This classification of digital games related to learning or "serious" purposes is still debated as stated in section 1.2.2. In which category should a game like Kerbal Space Program (KSP) be classified? KSP, as introduced in the synopsis, is a COTS

<sup>14</sup> <http://minutelabs.io/>

<sup>15</sup> <https://www.youtube.com/user/minutephysics>; 4.4 million subscribers as of 2018-08

<sup>16</sup> <https://www.youtube.com/minuteearth>; 1.9 million subscribers as of 2018-08

game based on realistic rocketry. It is realistic and therefore not a thematic game. It is not an educational game either as the first purpose of the project is profit. The closest category would be "popularization game" or "simulator" but the answer is not definite, as the game was not developed by professional rocket scientists. At any rate, KSP can be used in Serious Gaming, even more so when using mods that enhance or add realistic aspects to the game, by for instance using the names and sizes of real planets of the solar system.

The final part of this chapter deals with the assessment of learning and the demonstrated effects of digital game-based learning.

## 1.3 Learning strategies, assessment, and outcomes

### 1.3.1 Learning strategies

The expression *chocolate-covered broccoli* (Amy Bruckman, 1999; Laurel, 2002) depicts failed educational games which feature gaming sequences as extrinsic motivator (the chocolate) interspersed with learning and quizzes sequences (the broccoli). They failed because they failed to engage many of their users: they had a lower appeal compared to COTS games (section 1.2). Extrinsic motivation has now been rejected in most of the literature with a variety of approaches which try and protect the state of flow the player experiences. One such approach of using intrinsic motivation is stealth learning (Paras and Bizzocchi, 2005; Sharp, 2012): learning is maximal if users do not even realize that they are learning, which is akin to the idea that "kids like all humans love to learn when it isn't forced upon them" (Prensky, 2003). One way of implementing this strategy is to adopt COTS games codes and practices while providing "serious" content: COTS games are indeed very efficient at making their users able to play, they are educational games of their own mechanics and content. COTS games have excelled at "educating" their users for years, and this is demonstrated by their success in the last decades, transforming a niche entertainment into a mainstream, mass-market entertainment industry. *Cellcraft* presented in section 1.2.2 is an example of educational game that features intrinsic motivation and stealth learning: the learning is integrated into the game mechanics, and the game is not presented as a tool for education. Stealth learning is also the strategy used in *Hero.Coli* to teach synthetic biology.

### 1.3.2 Learning assessment

In order to assess the effects of game-based learning, several methods have been developed. For instance, authors have proposed three axes of analysis and evaluation of digital learning environments: Usability, Usefulness and Acceptability (Tricot et al., 2003). These three axes can also be applied to digital game-based learning as well. *Usefulness* measures the contribution that a video game can pro-

duce, that is to say, how much the player will learn. *Usability* measures the ease with which the player can perceive and interact with the game in order to achieve their goal. *Acceptability* measures the willingness of the player to use the game. Other methods of assessment rely extensively on the use of quizzes, multiple-choice questionnaires, other methods rely on game analytics and trace theory, or on combinations of those.

Quizzes are sequences in computer programs or games which prompt users to answer a series of questions. In the typical computer quiz, questions are presented one by one to the user, and must be answered by either clicking on the answer, pressing the button associated with the answer (A/B/C/D or 1/2/3/4 or specific gaming console buttons), or typing the answer in a text field. In rarer cases, these answers must be drawn or spoken into a connected microphone.

Quizzes are a controversial yet practical way to assess the level of a student because they match common academic practices: there will not be any problem on the student side to use computer quizzes, even though their engagement may be low and the metrics gathered hard to match to a skill or a know-how. In the case of multiple-choice questionnaires, it is also very easy to assess a whole class fairly using computer quizzes: answers can be defined unambiguously as wrong or correct, points can be attributed to each answer, computers do not discriminate students, and the process can be automatized. This is why the pretest-posttest design (Dimitrov and Rumrill, 2003) relies on multiple-choice questionnaires. Multiple-choice questionnaires are also highly reproducible, making it possible to do time-based comparisons using pretest-posttest studies, or cohort based-comparisons.

It has also been shown that quizzes, more than just assessing, can also have an educational impact: the assessment feedback is both assessment and learning strategy (Valerie J. Shute, Hansen, and Almond, 2008; Kleij et al., 2012). Outside the class, on the other hand, quizzes are only engaging to those who had already endeavored to learning. For instance, gamified learning apps described in section 1.2.4 can lead to great learning successes in non compulsory use. Otherwise, games integrating quizzes, due to their lower immersion and appeal cannot compete with COTS games.

Stealth assessment goes a step further than stealth learning by taking into account the risk of ruining the gaming experience by integrating assessment as described in section 1.3.1. As multiple-choice test assessments may transform the experience into *chocolate-covered broccoli* by breaking the flow of the game (Valerie J Shute, 2011), they should be avoided when possible. That is why research has been conducted on the analysis of interactions of players with the game and with other players: trace theory (É. Sanchez, Ney, and Labat, 2011; Clauzel, Sehaba, and Prié, 2011; Thomas et al., 2012; P. Bouvier et al., 2013; Patrice Bouvier, Lavoué, et al., 2013). Trace theory is based on the creation of player models constructed from iterative levels of interpretation of atomic observed elements or *obsels*. Obsels are simple, single events, upon which actions can be interpreted, upon which strategies and whole user paths can be reconstructed and analyzed through statistical analysis of data. These data can also be complemented with

data gathered from questionnaires and observations. This whole area of study called learning analytics has been active for years (Serrano-Laguna et al., 2012; Greller and Drachsler, 2012; Westera, Nadolski, and Hummel, 2014; É. Sanchez, Martinez-Emin, and Mandran, 2015).

Thanks to statistical analysis and advanced learning analytics, numerous studies demonstrate the possible outcomes that games can produce. These studies show however that these outcomes are possible, not that they are systematically present in all digital game-based learning implementations and uses. These outcomes are described in the next section.

### 1.3.3 Learning outcomes and issues

This section will only describe outcomes which are relevant to this thesis: learning and motivation. Indeed, Hero.Coli, the video game used as basis for this thesis aims at supporting learning and motivation in synthetic biology, there is no additional aim of behavioral or social effect.

Digital game-based learning has since proven its effects in multiple studies (Papastergiou, 2009; Annetta et al., 2009; Connolly et al., 2012; Girard, Ecalle, and Magnan, 2013; Granic, Lobel, and Engels, 2014; E. A. Boyle et al., 2016), comprising outcomes on knowledge acquisition, social, cognitive, motor skills as well as affective, motivational, behavioral outcomes. As an educational tool, the most interesting characteristics are engagement, as engagement is correlated with learning due to a fuller game experience and involvement, and curiosity, which may drive further endeavors. Both educational and popularization games have characteristics that make them efficient tools for learning and for raising interest. “The use of educational games within learning environments raises motivation, increases interest in the subject matter, intensifies information retention, encourages collaboration, and improves problem-solving skills.” (Schneider and Jimenez, 2012).

Studies on digital game-based learning also pinpoint the conditions required to generate outcomes. Connolly et al. (2012) summarizes these conditions by citing a previous work: "modern theories of effective learning suggest that learning is most effective when it is active, experiential, situated, problem-based and provides immediate feedback (E. Boyle, Connolly, and Hailey, 2011)". The active aspect of learning, a key principle in the constructionism theory, is underlined in several studies : “a serious game environment can promote learning and motivation, providing it includes features that prompt learners to actively process the educational content.” (Erhel and Jamet, 2013). The User Experience (UX) approach to game development also focuses on raising engagement in its motivation, participation, and involvement meanings (Hodent, 2017) by relying on mechanisms of neuroscience to analyze the mental processes of the users of a game. For instance, a core principle of UX is self-consistency in the design of the interface and of the feedbacks to avoid cognitive dissonance which can threaten immersion and

engagement.

In addition, some studies listed in Connolly et al. (2012) also focus on specific mechanisms such as the *relearning process* (Davidovitch, Parush, and Shtub, 2008), how to foster reflection (Kiili, Ketamo, and Lainema, 2011), how to foster learning using the game format, feedback, and the learning styles (Cameron and Dwyer, 2005). In this last paper, the game format did not improve retention of information compared to formal instruction, which shows that digital game-based learning is indeed not systematically successful.

In the field of cell biology in particular, several studies have already proven the impact VGs can have on learning and motivation (Stegman, 2014; Annetta et al., 2009).

Adaptive learning is another promising mechanism being studied. A program featuring adaptive learning can adjust its level of difficulty depending on the user's behavior: it can be more or less progressive - by adding, repeating, skipping steps in a step-by-step process; it can be more or less difficult by changing the complexity and speed of quizzes. It can be coupled with a system that lists knowledge items comprising a learning sequence under whatever form it is provided to the user - book, MOOC, video, game, ... In this system, a network of prerequisites and outcomes connect together the learning sequences, making it possible to suggest content to learners, whether they seem to be struggling due to shortcomings, or to be interested in a topic if they have learned several "chapters" related to a topic. The field of adaptive learning is still active and not yet widespread, as it is technologically difficult to generate content supporting these various levels of difficulty and these networks of knowledge. Research and applications are being tested in the US in Knewton<sup>17</sup>, and in Europe in DOMOSCIO<sup>18</sup>. Adaptive learning can be implemented as such in games by explicitly giving tutorials and feedbacks to the player when needed. It can also be adapted as an adaptive difficulty level, by evaluating in real time how much the user struggles in the game. Dynamic Game Difficulty Balancing and procedural generation of levels or enemies (Hunicke, 2005) as well as Adaptive Gameplay (Francillette, Gouaich, and Abrouk, 2017) are exactly that. Other experimental games using virtual reality also try and adapt to user experience, like the horror video game *Bring to Light*<sup>19</sup> which uses heart rate monitoring to evaluate the stress level of the player.

The proven effects of game-based learning and the tools needed to demonstrate them described in this chapter will be used in the next chapters of this thesis to determine whether the Hero.Coli video game can foster the learning of synthetic biology, which will be presented in the next chapter.

---

<sup>17</sup> Knewton is a private learning platform creating tailored courses for companies and universities. Website: <https://www.knewton.com/>

<sup>18</sup> <https://domoscio.com/>

<sup>19</sup> *Bring to Light*, Red Meat Games, 2018. Game page on the Steam game download platform: [https://store.steampowered.com/app/636720/Bring\\_to\\_Light/](https://store.steampowered.com/app/636720/Bring_to_Light/).



# Chapter 2

## Synthetic biology

This chapter describes the topic of Hero.Coli, the educational video game that will be presented in chapter 4. As discussed in the synopsis Hero.Coli was designed as a dissemination tool for Synthetic Biology (SB). Closely related to genetic engineering, SB is a relatively recent field developing upon previously developed techniques, methods, and concepts from other fields. It mainly draws from, interacts with, builds upon, and supports the fields of biochemistry, molecular biology, computational biology, systems biology, and engineering.

### 2.1 Definition

There have been many definitions given to SB: "Synthetic biologists engineer complex artificial biological systems to investigate natural biological phenomena and for a variety of applications" (Andrianantoandro et al., 2006), "SB is a research field that combines the investigative nature of biology with the constructive nature of engineering" (Purnick and Weiss, 2009). The shortest definition sums up SB to "an engineering technology based on living systems" (T. F. Knight, 2005). All definitions generally include:

- the use of engineering principles;
- artificial, synthetic biological systems as hardware, also known as wetware;
- research and non-research applications.

One characteristic that is often highlighted is *interdisciplinarity*. Namely, SB is considered interdisciplinary as it uses knowledge, practices, and methods from cell biology, molecular biology, systems biology, computational biology, and engineering among other fields.

Cell biology and molecular biology historically predate, and are the very basis of the fields SB draws from, as they describe their core elements: the cell, its chemical components, and their interactions. Theoretical SB relies on systems biology and computational biology for modeling and simulation purposes. Applied SB,

since it intends on synthesizing parts of or complete organisms, relies on practices developed for cell and molecular biology. These practices include cell culture, DNA extraction, polymerase chain reaction (PCR), sequencing, whole genome amplification, cloning, gene editing, insertion, deletion, microfluidics - high-throughput screening and selection of single cells - among others. Finally, applied SB also relies on engineering by borrowing from its practices and methods, such as the iterative process of design, implementation, and testing, as well as the use of well-defined and quantified standards. *Genetic engineering*, engineering methods to edit DNA, is therefore a subset of SB practice. The *engineering* theme in the definition of SB also implies that these biological systems have defined and controlled characteristics.

SB relies heavily on the *production of proteins* by the cell to achieve useful functions. This fundamental phenomenon is described in the next section for the purpose of clarity in further sections.

### 2.1.1 The Central Dogma

The central dogma of molecular biology explains the protein synthesis (protein production) process by the cell. Proteins are huge molecules, composed of long strings of amino-acids containing numerous atoms: for example, in *E. coli* they comprise an average of 300 up to 2367 amino-acids according to Bionumbers, the database maintained by Harvard University<sup>1</sup>. Each amino-acid containing at least a dozen atoms, the average size of *E. coli* proteins is 35 kDa, *i.e.* an average 35K atoms per protein<sup>2</sup>. Proteins are essential to the cell because they are the main molecules that carry out functions - they act inside the cell but also outside of it, on its environment - they form the cell membrane and internal cell structures, are involved in metabolism, cell repair, DNA replication. Some of the most well-known and studied proteins are:

- the Green Fluorescent Protein (GFP): a protein frequently used in SB as a marker because it fluoresces in green when receiving blue light;
- keratin: the main component of hair and fingernails;
- insulin: a protein which regulates our blood glucose level;
- hemoglobin: a protein which fixes oxygen into our red blood cells;
- collagen: a structural protein;
- membrane transport proteins: a family of specialized proteins which transport specific molecules across cell membranes.

---

<sup>1</sup> Last access 2018-09: <http://bionumbers.hms.harvard.edu/bionumber.aspx?s=n&v=3&id=108985>

<sup>2</sup> Last access 2018-09: <http://bionumbers.hms.harvard.edu/bionumber.aspx?id=113349&ver=0&trm=protein&org=>

Proteins are not the only molecules that the cell uses to live and grow - simpler and smaller molecules, and DNA and RNA are also directly involved - but they comprise the bulk of those vital molecules.

This pivotal phenomenon of protein production can be understood through the Central Dogma of molecular biology. The Central Dogma framework describes how genetic information carried by DNA is *transcribed* into RNA which is then in turn *translated* into proteins - cf. figure 2.1.

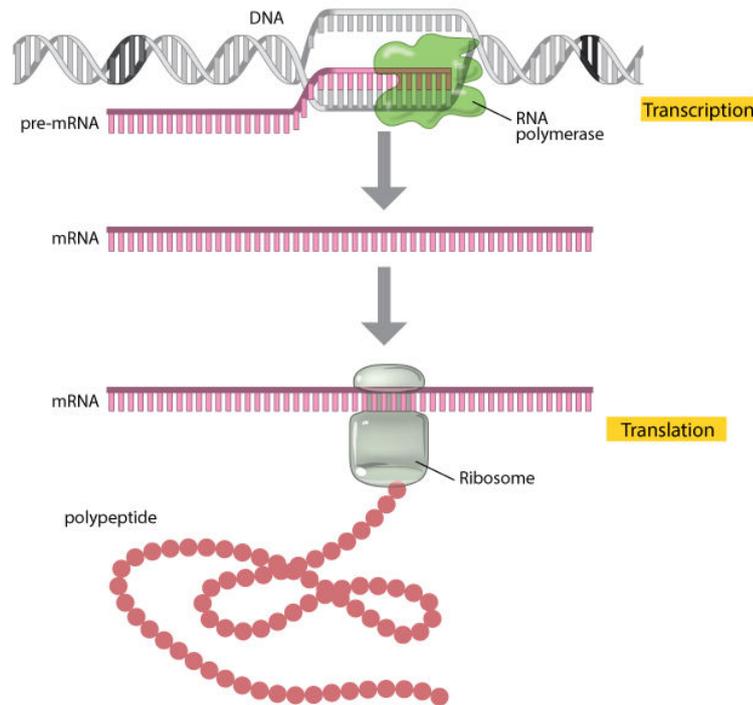


Figure 2.1: Consequences of the Central Dogma of molecular biology: transcription and translation

Source: *Nature*, 2013 <https://www.nature.com/scitable/topicpage/translation-dna-to-mrna-to-protein-393>

Transcription is the process of creation of a single-strand RNA molecule from a double-strand DNA molecule. The information encoded in the DNA "alphabet" ACGT (the four nucleobases adenine, cytosine, guanine, thymine) to the RNA "alphabet" ACGU (the same first three nucleobases adenine, cytosine, and guanine, and uracil instead of thymine) into a new RNA strand by the *RNA polymerase*. During translation, RNA nucleobases are read and translated triplet by triplet into a chain of amino-acids by *ribosomes*. This chain is called a peptide, a polypeptide, or a protein depending on its length and on the context, but it is more of a vocabulary distinction than one based on functionality. For simplicity, these chains will be referred to as proteins.

As we will see later, by knowing the characteristics of the DNA sequence in-

volved in the production of a given protein, it is possible to model the production of this protein. Having presented how cells produce proteins, the core principles of SB will be described regarding the goals of SB and constraints imposed by protein production.

## 2.2 Principles

The general goals and means of SB are, to summarize, designing the genome of new organisms to solve a problem by the action of the proteins they produce. From these goals and means, Drew Endy listed a set of principles that should guide further developments in SB. In his seminal paper called “Foundations for engineering biology” (Endy, 2005) Drew Endy underlined the main process future SB projects should follow: designing biological circuits on paper in the same manner as boolean circuitry, so that SB design would be as easy as electronics. He listed 3 core principles which seemed primordial to him in order to reach that goal - and which have been indeed built upon ever since: *Decoupling*, *Standardization*, and *Abstraction*. These three core principles are to be the basic content of any learning process of SB, including Hero.Coli in chapter 4.

### 2.2.1 Decoupling

Electronic components can be assembled easily because, although they change the overall behavior of the circuit they are put into, they do not change the behavior of the other individual electrical components already present in the circuit. These components are *decoupled* - one component does not change the behavior of another component. Decoupling in SB is similar. It is making the assumption that, at first approximation, genetic sequences lastly added to a system will not modify the behavior of previously present genetic sequences. In order to be able to follow that assumption in practice, synthetic biologists design genetic elements so that they are not context-dependent.

### 2.2.2 Standardization

The units of physical measurement of the international system, the SI units, have helped physicists and chemists build a worldwide common theory and practice. Constants and physical properties - such as the mass and charge of particles - are now measured or computed in these units, making it easier for teams around the world to collaborate. Similarly, internationally established common languages and standards for SB would be greatly beneficial. Several such standards have already been proposed and implemented such as BioBricks<sup>TM</sup> (easy-assembly DNA sequences), cell chassis (well-characterized cells to work with), and descriptive languages like the Synthetic Biology Open Language (SBOL).

Standardization lead to the creation of BioBricks, which will be further described along with the principles behind their creation and the specific bricks used

in *Hero.Coli* in their biological context for later reference.

## BioBricks

A restrictive approach of practicing SB is the creation of biological devices using standardized, well-characterized sequences of DNA. There are different standards to define these sequences. One of the most widespread are the BioBricks created by Knight *et al.* at MIT (T. Knight, 2003). The BioBricks Foundation is the main institution advocating for their use<sup>3</sup>. BioBricks were designed to be interchangeable, just like Lego bricks, so that their assembly itself would not be a challenge each time they would be used. Establishing this standard made it easier to design genetic systems, to assemble them, to exchange them internationally, and to characterize them in a digital library. This library is the Registry of Standard Biological Parts<sup>4</sup>. The thousands of bricks can be browsed by type, function, context of creation, cell chassis, and other criteria. Some of the functions of the bricks are shown in figure 2.2.

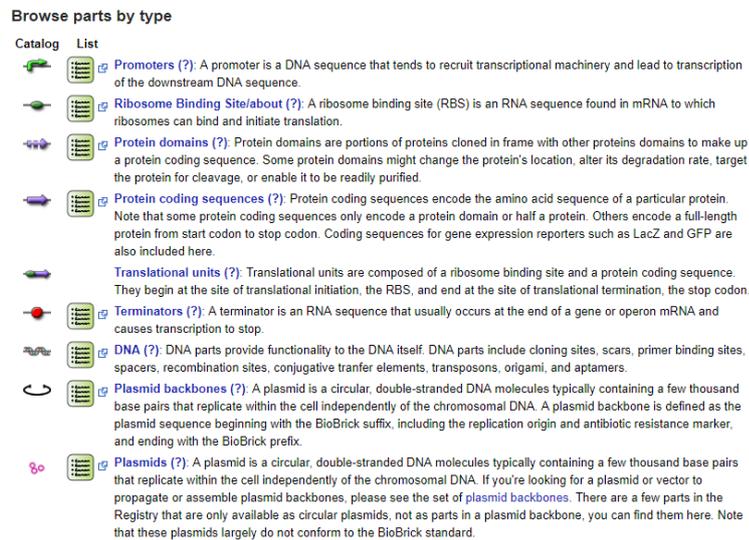


Figure 2.2: Registry of Standard Biological Parts: brick functions

Source: website of the Registry of Standard Biological Parts <http://parts.igem.org/Catalog>

Among the most common ones are *promoters*, *ribosome binding sites* (RBSs), *coding sequences* (CDSs), and *terminators*. These function names are actually the names of DNA regions implied in the process of gene expression presented in section 2.1.1:

- promoters recruit RNA polymerase and therefore mark the start of the DNA sequence to be transcribed;

<sup>3</sup> Founded in 2006. <https://biobricks.org/>

<sup>4</sup> Founded in 2003. <http://parts.igem.org>

- transcribed RBSs recruit ribosomes and therefore mark the start of the RNA sequence to be translated;
- CDSs are DNA regions coding for a protein;
- terminators cause the RNA polymerase to stop transcribing.

The role of BioBricks in transcription is shown in figure 2.3.

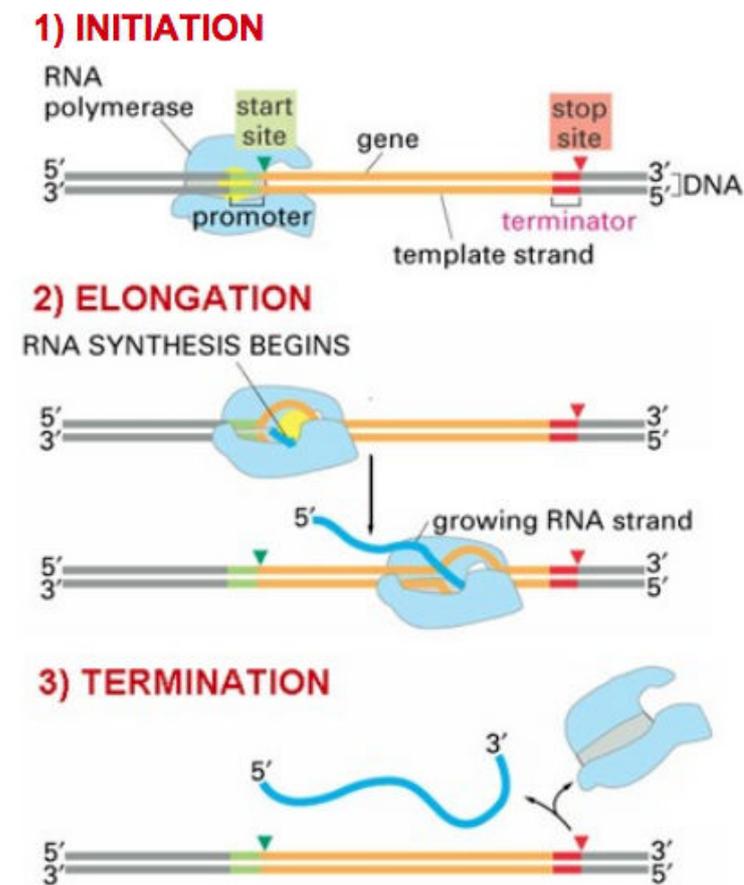


Figure 2.3: BioBricks in the transcription process

Source: Joan Cañete <https://www.slideshare.net/smilewiththisgirl/dna-replicationtranscription-and-translation>

The next sections detail the mechanisms governing each of those four types of bricks, as they are the bricks used in Hero.Coli.

**Promoters** Promoters regulate when the downstream DNA sequence will be expressed, by allowing the RNA polymerase to bind to the DNA. Some promoters, the *constitutive promoters*, have the downstream sequence always expressed. *Inducible* and *repressible* promoters are respectively active or inactive in the presence of a compound, respectively called an activator or a repressor. For instance, the pBad promoter is active in the presence of the arabinose sugar, so we consider arabinose to be the activator of the pBad promoter. This means that a DNA strand containing a sequence coding for a protein appended to a sequence containing pBad will be transcribed to RNA only when arabinose is present, and due to the relatively short lifetime of RNA<sup>5</sup>, the protein coded in the RNA will be expressed only when arabinose is present. Conversely, lactose, another sugar, is the repressor of the pLac promoter and the pLac promoter is only active in the absence of lactose.

However, these promoters do not behave like logic gates: they do not perfectly switch from active to inactive state. For instance, inducible promoters do not guarantee a 100% expression rate above and 0% expression below a particular threshold of inducer concentration. The expression of an inducible system follows the Hill equation<sup>6</sup>, which follows a sigmoid curve when plotted against the concentration of inducer, as shown in figure 2.4 (Siegele and Hu, 1997).

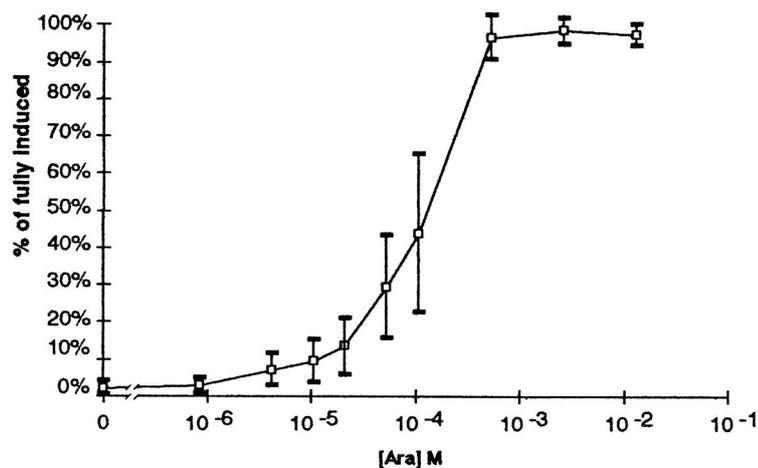


Figure 2.4: The Hill function, typical gene expression from an inducible system  
Source: Siegele and Hu (1997)

Different promoters have different concentration thresholds and different *leakinesses*, *i.e.* asymptotes and growth rates (slope near the threshold concentration). Some bricks have similar functions but are quantitatively different - they increase

<sup>5</sup> "The lifetime of mRNA molecules is usually short in comparison with the fundamental time scale of cell biology defined by the time between cell divisions. [...]For *E. coli*, the majority of mRNA molecules have lifetimes between 3 and 8 minutes.". Source: <http://book.bionumbers.org/how-fast-do-rnas-and-proteins-degrade/>, last access 2018-09.

<sup>6</sup> Mathematical details are given in section 2.2.3.

the expression of the downstream protein, or are more influenced by external parameters. This is also true for other types of BioBricks: some bricks are only quantitatively different from other bricks of the same type.

**Ribosome Binding Sites (RBSs)** Also called ribosomal binding sites, RBSs are DNA sequences which, once transcribed to RNA, bind to ribosomes. "Strong" RBSs will be transcribed to RNA sequences which will tend to more easily bind to ribosomes than "weak" ones, whose RNA sequences will bind less easily. Therefore, RNA regions downstream a sequence originating from a strong RBS will be more accessed by ribosomes: any protein encoded there will be more translated, resulting in greater protein expression. This RBS "strength" is formally called the *level of expression* of the RBS.

**Coding Sequences (CDSs)** Also called coding regions, CDSs are DNA sequences which, when transcribed to RNA, code for a protein. The protein assembly phase of the translation process happens when the ribosome traverses the coding sequence.

**Terminators** Terminators are DNA sequences which make the RNA polymerase unbind, detach from the DNA sequence. They therefore cause the end of the transcription. Different terminators in different contexts may cause unbindings with different efficiencies: terminators of high efficiency will cause the RNA polymerase to unbind most of the time while low-efficiency terminators cause unbinding only rarely. In Reynolds, Bermúdez-Cruz, and Chamberlin (1992), termination efficiency ranges from 2% to 90%. This makes it possible to control the expression of successive promoter-RBS-CDS-terminator sequences: a high-efficiency will make the expression of two successive sequences uncorrelated, while a low efficiency terminator will increase the transcription of the downstream sequence in proportion to the transcription of the upstream sequence. However, this mechanism of control of expression through terminator efficiency is rather complex, and we decided not to include it in Hero.Coli for clarity purposes. It is included here as a side note to justify the existence of several terminators. The only terminator available in the game is the double terminator, of measured efficiency between 0.97 and 0.984<sup>7</sup>, rounded to 1 in the game.

Having introduced these four types of BioBricks, the very process by which they are being created will be presented. Thanks to the catalyzing effect of an international competition, the standardization effort of creating BioBricks is supported by a whole community of synthetic biologists, using open source and open data tools.

---

<sup>7</sup> Source: [http://parts.igem.org/Part:BBa\\_B0015](http://parts.igem.org/Part:BBa_B0015).

## iGEM

The bricks presented in the previous section were submitted by contributors, among whom many participants in the International Genetically Engineered Machine (iGEM) competition. The iGEM and BioBricks foundations work closely together to create and stimulate an international community of synthetic biologists. The iGEM competition and its associated foundation are "dedicated to education and competition, the advancement of SB, and the development of an open community and collaboration"<sup>8</sup>. Participants present their work to the whole community and compete in different categories - high school, undergraduate, and "overgraduate". They use SB, and more precisely the BioBricks, to propose interesting and potentially useful designs of new, synthetic organisms. For instance, in 2017 the TUDelft team, "overgraduate" Grand Prize winners, proposed to tackle antibiotic resistance<sup>9</sup>. The Vilnius-Lithuania team, undergraduate Grand Prize winners, developed a framework to facilitate multi-plasmid projects in SB<sup>10</sup> (plasmids are circular strands of DNA found in bacteria, that are used in SB as carriers to inject DNA into cells). Each year since 2004, teams from all over the world have gathered in Boston; in 2017, there were "nearly 5,400 participants, from 310 teams, representing 44 countries"<sup>11</sup>. It must be noted that in an effort to promote open-source and collaboration, compatibility between BioBricks and the Synthetic Biology Open Language<sup>12</sup> (SBOL) is supported by iGEM and the BioBricks Foundation<sup>13</sup>. SBOL is an open language used by synthetic biologists to specify genetic constructs. It uses the same type of icons in its component SBOL Visual than in iGEM's Registry of Standard Biological Parts (see figure 2.2 and figure 2.5), and in its SBOL Data component proposes a data format to represent constructs.

Standardization has made collaboration and the building of the SB community easier. In order to increase the predictability of and control over the new organisms being synthesized, synthetic biologists also have to work on abstracting the biological systems.

### 2.2.3 Abstraction

Electronic devices are composed of a complex assembly of materials with specific physical properties, obeying multiple physical laws - Van der Waals forces, Maxwell's electromagnetic laws, quantum effects, thermal dilatation, and vibration among others. However, in the daily conditions of use, at human scale, these devices obey a smaller, simpler and less fundamental set of laws and their behavior can be predicted and approximated to simpler equations. Resistors obey Ohm's law, silicon semiconductors act as switches. The behavior of electronic circuits

---

<sup>8</sup> <http://igem.org/About>

<sup>9</sup> <http://2017.igem.org/Team:TUDelft/Description>

<sup>10</sup> <http://2017.igem.org/Team:Vilnius-Lithuania>

<sup>11</sup> Source: <https://after.igem.org/news/29124>.

<sup>12</sup> <http://sbolstandard.org>

<sup>13</sup> [http://2018.igem.org/Resources/Software\\_Tools](http://2018.igem.org/Resources/Software_Tools)

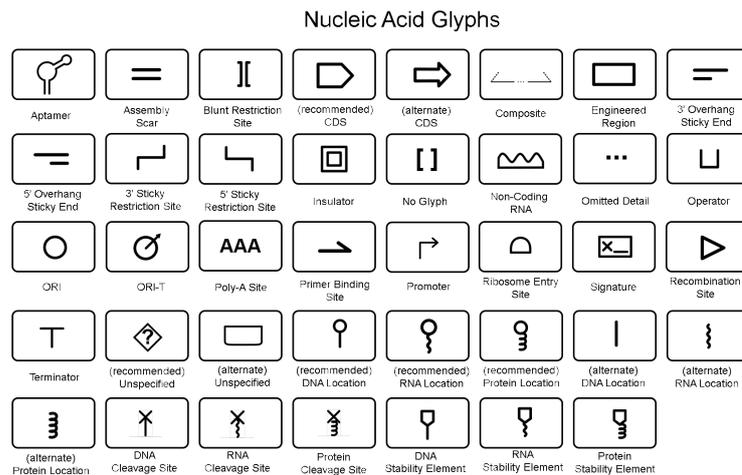


Figure 2.5: SBOL Visual: glyphs representing functional DNA sequences  
 Source: website of SBOL Visual <http://sbolstandard.org/visual/glyphs/>

can be computed from the behavior of their parts, and the characteristics of, for instance, electronic oscillators can be predicted. Similarly, through abstraction, SB aims at modeling living systems using maths and boolean logic. With some assumptions and approximations - simplifying the context of use, assuming the constance of some parameters, neglecting some phenomena - the behavior of biological circuitry can be modeled by equations. This modeling is accomplished through computational and systems biology, two closely intertwined disciplines. Computational biology is basically the use of computer science techniques to model and simulate any biological object, and systems biology is the use of graph theory and maths to model and simulate complex biological phenomena on the scale of systems.

## Computational biology

Computational biology is also called *in silico* or *in numero* biology by analogy with *in vitro* and *in vivo*. It is also sometimes improperly referred to as bioinformatics. Bioinformatics is usually restricted to data storing and processing applied to biology (Murphy, 2018). An example of computer-based processing of biological data is DNA sequence alignment.

The core principle of computational biology is to use computer science tools to study biological phenomena. Mathematical modeling and computer-based simulation, but also data analysis are its main components. The main inputs from computational biology to SB are the modeling and simulation on computers of biochemical reactions, using for instance the law of mass action (equilibrium of reaction), Michaelis-Menten kinetics (speed of reaction), enzymatic reaction models, and cell models in general. Other inputs are linked to the modeling of evolutionary processes and to DNA sequence matching.

## Systems biology

"Systems biology is the study of biological systems whose behaviour cannot be reduced to the linear sum of their parts' functions" (*Systems biology - Latest research and news* / *Nature* 2018). This field specializes in describing biological mechanisms as the result of the behavior of systems. Population dynamics, metabolic pathways, gene regulatory networks are examples of biological complex systems studied using systems biology. Other examples include applications from the *-omics* disciplines - proteomics, glycomics, epigenomics... Biological systems are studied in order to determine the global behavior and emergent properties arising from components and subsystems characteristics. Steady-states are investigated to determine whether they are stable or unstable equilibrium states, or oscillating states. Systems biology clearly overlaps with computational biology in today's applications, as computer-based simulations are the norm.

To give examples of modeling in biology, some models of phenomena pertaining to cell biology will be presented in the next section, including their biological description and their modeling in mathematical terms. This will not be just a list of handpicked examples: these phenomena are all included in the simulator of Hero.Coli.

## Models in cell biology

A range of cell biology phenomena can be modeled using mathematical equations. Protein expression, protein degradation, enzymatic reactions, osmosis, and active transport are among the phenomena simulated in Hero.Coli. The mathematical expressions will be only slightly adapted when ported from their differential equation form to their applied form in the code of Hero.Coli.

**Protein expression** The general process of protein expression is presented in paragraph 2.1.1. To present the modeling of protein expression, I will use the production of the Green Fluorescent Protein (GFP), a protein presented in the same earlier paragraph. GFP will be produced by a cell which contains the sequence of BioBricks shown in figure 2.6.

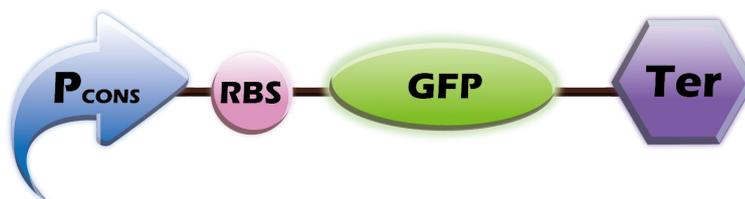


Figure 2.6: BioBrick sequence producing GFP.

Source: <http://parts.igem.org/File:GFP2015.png> by user SherryNCTU on 18 September 2015, last access 2018-08.

This sequence contains:

- a *constitutive promoter*  $P_{cons}$ ;
- an RBS;
- a GFP coding sequence;
- a double terminator.

From this sequence, the protein expression of the genetic device is computed as:

$$\frac{dGFP_{expr}(t)}{dt} = r_{Pr} * RBSf * Tf * \beta$$

with:

- $\frac{dGFP_{expr}(t)}{dt}$ : the change of amount of GFP due to the production of GFP, in  $mol.s^{-1}$  or number of molecules per second.
- $r_{Pr}$ : the dimensionless factor taking into account promoter regulation. In our simple case of a constitutive promoter:  $r_{Pr} = 1$ .

In the case of positively regulating inducible promoters, with  $I$  being the amount of inducer:

$$r_{Pr} = H(I, K_I, n_I)$$

where  $H(x, K, n)$  is the Hill function (a type of sigmoid), evoked in section 2.2.2, of expression

$$H(x, K, n) = \frac{x^n}{K + x^n}$$

In the case of negatively regulating promoters,

$$r_{Pr} = 1 - H(I, K_I, n_I)$$

With simultaneous regulations, the global regulation formula is the product of those Hill functions.

- $RBSf$ : the RBS factor, dimensionless, between 0 and 1, that corresponds to the RBS affinity with the ribosomes.
- $Tf$ : the terminator factor, dimensionless, between 0 and 1, representing terminator efficiency - see section 2.2.2 about terminator efficiency.
- $\beta$ : the maximal protein production rate, in  $mol.s^{-1}$  or number of molecules per second.

**Protein degradation** Degradation is the spontaneous breaking down of a molecule. This phenomenon leads to the progressive disappearance of a molecule that is not being constantly produced. Proteins degrade following the exponential law. It means that the change of quantity of a protein due to degradation can be modeled by the equation:

$$\frac{dP_{deg}(t)}{dt} = -\lambda P(t)$$

with:

- $\frac{dP_{deg}(t)}{dt}$ : the change of quantity of  $P$  due to degradation, in  $mol.s^{-1}$  or number of molecules per second.
- $P(t)$ : the amount of the protein at the time  $t$ , in  $mol$  or number of molecules.
- $\lambda$ : the degradation rate of this protein, in  $s^{-1}$ . It depends on the temperature.

**Enzymatic activity** Enzymes are large molecules which act as specific catalysts - they make a specific chemical reaction happen faster, and are not consumed during the reaction. The reaction would occur without the enzyme, but slower, potentially to the point where it would barely occur. In extreme cases, some reactions take millions of years to reach equilibrium without the enzyme, and milliseconds with it (Radzicka and Wolfenden, 1995). Put simply, enzymatic reactions are chemical reactions of the type:



where  $E$ , the *enzyme*, acts as catalyst and is thus present in the beginning and in the end,  $S$  is the *substrate*, which reacts to become one or several *products*  $P$ . The use of Michaelis-Menten kinetics gives the following equation for the reaction velocity<sup>14</sup>:

$$\frac{dP_{enz}(t)}{dt} = \frac{V_{max} * S}{K_m + S}$$

where:

- $\frac{dP_{enz}(t)}{dt}$ : the change of quantity of  $P$  due to the enzymatic reaction, in  $mol.s^{-1}$  or number of molecules per second.
- $V_{max}$ : the maximal product production rate, in  $mol.s^{-1}$  or number of molecules per second.
- $S$ : the amount of substrate, in  $mol$  or number of molecules.
- $K_m$ : a constant representing affinity between substrate  $S$  and enzyme  $E$ , in  $mol$  or number of molecules.

---

<sup>14</sup>Source, last accessed 2018-08: <http://2012.igem.org/Team:Wisconsin-Madison/modeling>

**Osmosis** Osmosis is, generally speaking, the tendency for chemical species to diffuse. The diffusion described here happens across the cell membrane. The diffusion of small chemical species across the cell membrane obeys Fick's first law<sup>15</sup>:

$$\frac{dC_{cell}}{dt} = (C_{ext} - C_{cell}) * \mu * S$$

with:

- $C_{cell}$ : the amount of the chemical species in the cell, in *mol* or number of molecules;
- $C_{ext}$ : the amount of the chemical species outside of the cell;
- $\mu$ : the permeability of the cell wall relatively to the chemical species. If the chemical species is too big - for instance if it is a protein - then  $\mu = 0$  because the species cannot cross the cell wall. Its dimension is *mol.s*<sup>-1</sup> or *s*<sup>-1</sup>.
- $S$ : the contact surface between the cell and the medium containing the chemical species. If the cell is completely submerged in the medium,  $S$  is the surface of the cell. Expressed in *m*<sup>2</sup>.

**Active transport** Cells can actively force some molecules to cross the cell membrane in an energy-consuming process, against the spontaneous phenomenon of osmosis, to build up a concentration. This happens, for example, when neurons build up neurotransmitters before releasing them in synapses to transmit a neural influx. The mechanism is referred to as active transport. The cellular elements involved in active transport are the membrane transport proteins, mentioned as an example of protein in section 2.1.1. For instance, specific membrane elements called ions channels force  $Na^+$  and  $Ca^{2+}$  ions into specialized sensory neurons in the presence of heat or capsaicin, a component of spicy peppers (Caterina et al., 1997).

All of those models of cell biology phenomena can be conflated into a *whole cell model*. Different types of cell models will be presented to justify the technical choice of implementation of the simulator of the Hero.Coli video game.

## Whole cell models

Whole cell models can be interesting to study a particular aspect of cells - their metabolism, their growth, their robustness when facing stress. They can also be used to predict the behavior of a future synthetic organism still under development.

Cell models can become a pivotal tool in SB as they theoretically allow for the early detection of consequences of DNA editing. A comprehensive enough model

---

<sup>15</sup> Fick's first law on Wikiversity: [https://en.wikiversity.org/wiki/Diffusion\\_and\\_osmosis](https://en.wikiversity.org/wiki/Diffusion_and_osmosis). Last access 2018-09.

can help detect precise interference issues in metabolic pathways, while coarse-grained models help discover broader problems, such as ribosomal depletion or saturation. Among the most comprehensive models are the models developed at the Karr Lab. They build on the extensive - yet still not quite complete - knowledge of the cell machinery, cumulating different techniques to model different mechanisms. The *E. coli* model developed at the Karr Lab (Markus W. Covert et al., 2008) uses "a flux balance analysis model of global metabolism, an ODE [ordinary differential equation] model of central carbon metabolism, and a Boolean model of transcriptional regulation" (source: <http://www.karrlab.org/research>). Their whole-cell computational model of *Mycoplasma genitalium* (Karr et al., 2012) allowed for the simulation and monitoring of the entire life cycle of this bacterium, while also developing computational techniques to allow for the simulation of the complete metabolism (see figure 2.7).

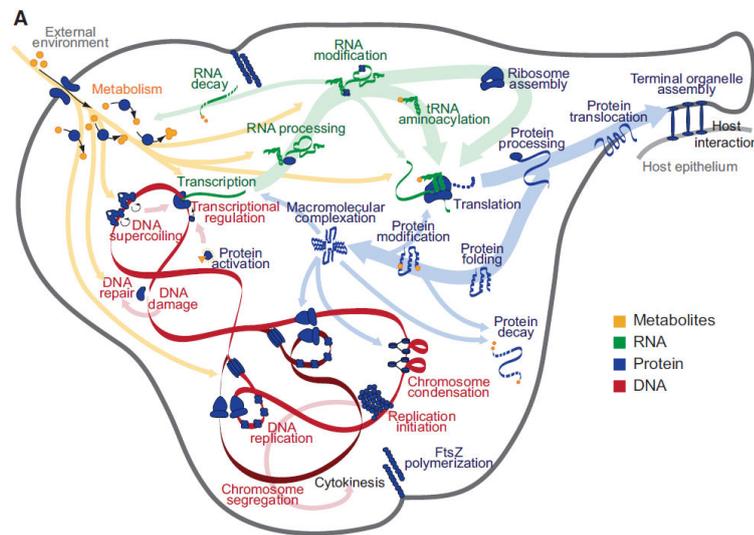


Figure 2.7: Cell mechanisms modeled in Karr's model  
Source: Karr et al. (2012)

While these models offer valuable insights into cells mechanisms and the future of SB, they also are very complex and long to build. The *Mycoplasma genitalium* model (Karr et al., 2012) required reviewing 900 publications in order to gather the values of more than 1900 parameters. Such whole-cell models can also be computationally demanding and thus may be impossible to use for real-time simulations, or for the modeling of cell colonies.

Coarse-grained models, on the other hand, allow for quicker iteration in their development process of adjusting parameters, while still producing valuable insights. They do not model in detail all the sub-mechanisms, but make simplifying assumptions about the state of the cell, or its environment. This way, far less sub-components and associated parameters are included (Teusink et al., 2000; Weiß et al., 2015), allowing for easier model building, tuning, and computational ex-

ecution. Examples of two such coarse grain models are given in figure 2.8 and figure 2.9.

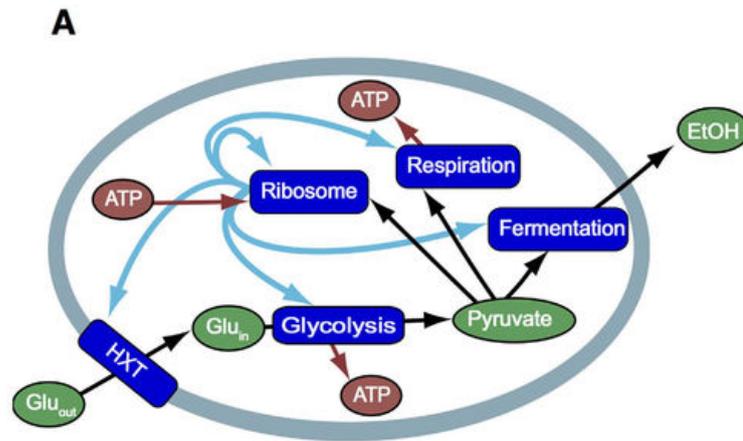


Figure 2.8: Model developed by Wortel et al.  
*Source: Wortel et al. (2016)*

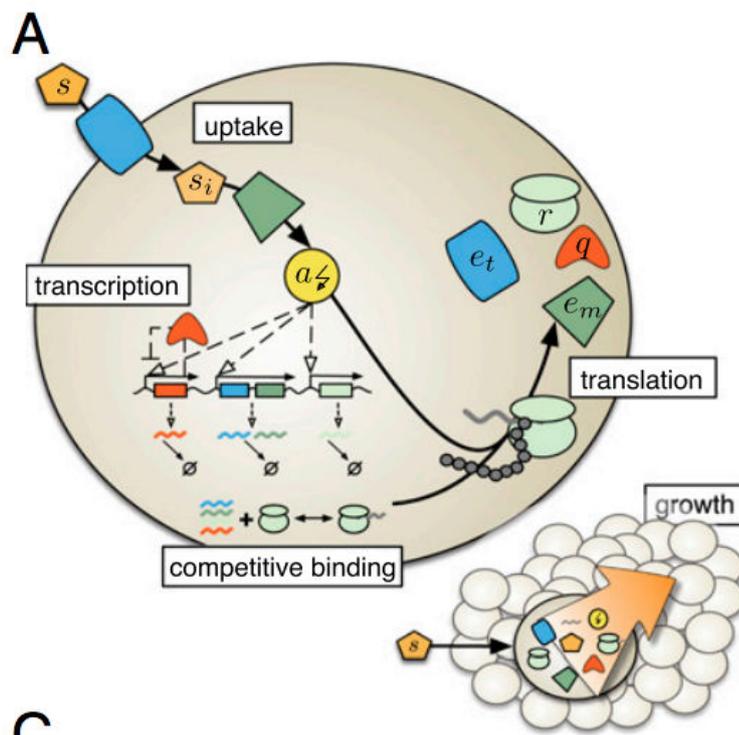


Figure 2.9: Model developed by Weiße et al.  
 Source: Weiße et al. (2015)

Even being based on the principles of decoupling, standardization, and abstraction, BioBricks are still not perfectly predictable and interchangeable. The reason behind this is due to a series of limitations that are inherent to biological systems: complexity, variation, and evolution. These limitations are presented in the next section.

## 2.3 Limitations of synthetic biology

SB is limited by the very hardware it uses - wetware, biological matter. In his seminal paper cited in the previous section to present principles of SB (section 2.2), Endy (2005), Drew Endy also listed the problems that plagued the then nascent SB. He focused on the three hardest problems to cope with due to their inherent nature which will be briefly described here: complexity, variation, and evolution.

### 2.3.1 Complexity

Biology is a field characterized, or one may even say plagued by complexity, as has become increasingly apparent from the work in several *-omics* sub-fields. For example, we know that metabolic pathways form intricate networks of interactions, several biological elements are implied in different parallel systems: energy production, information storage, signaling are linked through the molecule adenine, which is a nucleobase but also a precursor of adenosine and derivatives, involved in energy transfer as adenosine triphosphate (ATP), and signal transduction of many biological processes as cyclic adenosine monophosphate (cAMP). This complexity affects SB in two ways: it makes the chemical synthesis of genetic systems difficult, and it also makes decoupling hard to achieve. Several DNA mechanisms greatly hinder decoupling, including (1) natural DNA protection systems, and (2) direct interactions between proteins, RNA, and DNA especially in prokaryotes like *E. coli*. This makes it difficult to characterize BioBricks in isolation. Synthetic biologists can only create BioBricks with the smallest working context conditions, and list those conditions when characterizing the BioBricks.

### 2.3.2 Variation

Biological systems are inherently stochastic. For example, we know that molecules move according to Brownian motion. This motion, in itself random, depends on temperature, which can change according to heat transfer, as well as heat production or consumption due to chemical reactions. Many other biochemical phenomena are also stochastic, some of them described in section 2.2.3. Small genetic mutations and accidents within a population are random and can create small genetic variations with far fetching consequences. Even bacteria with the same genome will have different fates, effectively following different strategies in the same controlled conditions. For instance, some will grow and age more than others; in case of food scarcity, a proportion of bacteria will become *persisters*,

*i.e.* dormant, in stasis. Concentration gradients of important chemical species like nutrients can also introduce stochastic variation. The lack of understanding of inherent noise in living systems leads to unreliability in design (Elowitz, Levine, et al., 2002). Cell modeling at the scale of colonies is usually reliable, but is very hard to achieve at the unicellular level.

### 2.3.3 Evolution

Evolution in itself is a huge obstacle for SB. Synthetic genetic devices and synthetic organisms may be suboptimal, therefore evolution may get rid of those synthetic genetic devices. For instance, genetic devices constructed using SB techniques and introduced in the host organisms may be energetically demanding, or may make the host organism less resilient to external stress. We know that host bacteria seem to search for a "solution of a cost–benefit optimization problem" through protein expression (Dekel and Alon, 2005), therefore, if this synthetic DNA is suboptimal, host bacteria will divide less, die more, or get rid of genetic vectors called *plasmids* used by synthetic biologists to integrate synthetic DNA. This means that, if the host bacteria are placed among *wild type* bacteria (bacteria that were not genetically modified), their synthetic genetic devices will not be selected for and will eventually disappear, because wild type bacteria will outcompete them. Even among a completely synthetic population, evolution will take place and may interfere. Random mutations and other genetic accidents will occur and may inactivate the synthetic genetic devices. These mutants, organisms with inactivated synthetic genetic device, may have higher fitness and thus could eventually outcompete and eliminate synthetic organisms from the population. This will happen even though bacteria do possess a natural mechanism which prevents the production of unneeded proteins from stalling growth (Shachrai et al., 2010). Other mechanisms such as *horizontal gene transfer* may interfere with the work of synthetic biologists: bacteria sometimes exchange genetic material, making it harder to control which cells have which genetic devices.

The goals, principles and issues of SB having been presented, the next section will list the most historically significant or typical uses and applications of SB. This will help define the stakes of teaching and popularizing SB.

## 2.4 Uses of synthetic biology

Uses of SB comprise theoretical and applied research, the production of chemicals in the pharmaceutical, energy, and food industries as well as a few others uses in biohacking, scientific popularization, and art.

The most iconic application in the field is the production of insulin by genetically modified bacteria in 1978 (Villa-Komaroff et al., 1978). More recent examples are the vitamin-A enriched Golden Rice (Ye et al., 2000) and the production of a precursor of artemisinin, a drug used in the treatment of malaria (Christopher J Paddon et al., 2013; Chris J. Paddon and Keasling, 2014). The authors have

integrated the metabolic pathway of production of artemisinic acid from the plant *Artemisia annua* into the *E. coli* bacterium in order to ease the process of growth and extraction of the active chemical compound. Other health-related applications are gene and engineered-cell therapies (Kitada et al., 2018). Instead of relying on genetically engineered organisms to produce a specific drug, and then having patients take the drug, engineered-cell therapies aim at injecting the patients with cells that will react and adopt a therapeutic behavior when symptoms are detected. Gene therapies aim at modifying the genome of patients to cure them from mono- and polygenic disorders - diseases caused by one or several genes involved. Usually, a neutralized virus is used as a vector to carry and implant DNA into the patient's DNA.

In research, the landmark papers for SB are technical proofs-of-concepts of a genetic toggle switch (Gardner, Cantor, and Collins, 2000) and of a genetic oscillator (Elowitz and Leibler, 2000). In another research project focusing on the study of genetic mechanisms, some researchers have synthesized two more nucleobases in order to understand why DNA uses a four-base code (ATGC) (Malyshev et al., 2014).

GMOs used in the food industry are also very famous, due to their controversial nature. On the one hand, they hold promises of higher yields, higher nutritious content, better taste, longer periods of consumption (products stay fresh longer), resistance to parasites, diseases, drought, frost, salinity. On the other hand, there are threats of accidental gene transfer, be it between two closely related groups of organisms (Jr, Halfhill, and Warwick, 2003), or between two different, distant species through horizontal gene transfer (HGT) (Keese, 2008). To quote this paper, "[HGT] to humans has been controversially proposed as a potential trigger for oncogenesis". Additionally, HGT of herbicide-resistance genes to wild non-cultivated plants would nullify the use of herbicides. There are also threats of uncontrolled GMO spread due to contamination through the wind or by accidental transport of seeds or pollen. This uncontrolled GMO spread could endanger traditional or endemic varieties of crops and lead to human dependency on a small number of cultivated species. Relying on a very small number of crops with low genetic variability is very dangerous: it makes it easier for diseases to spread, and makes it harder for replacement crops to be quickly available. For instance, it led to the Great Famine in Ireland. Lasting from 1845 to 1852, it killed one million people, and forced two million people to exile. Overly dependent on the Irish Lumper variety of potatoes, the Irish population lost its main source of food when blight spread quickly across the country. Low crop variety might also lead to the extinction of the over-cultivated cavendish variety of bananas (Pearce, 2003). Practices such as the production of sterile plants and clone plants (to have a perfectly controlled plant) increase these risks, in addition to putting more financial stress onto farmers: they cannot reuse their own seeds and breed selectively to create new varieties. Seed patenting by some companies creating GMOs have the same effects. This explains why SB appears as a double-edged sword, in addition to yet unsubstantiated fears that GMOs would be inherently dangerous - this is

further discussed below, in section 2.5.1.

Another striking example of SB application belongs to the energy sector: the production of biofuels (S. K. Lee et al., 2008; Liu and Khosla, 2010). Other applications include industrial enzymes, biological computers and data storage in DNA, biosensors, new materials and new 3d printing processes (Gentry, 2013).

In all of those examples and in SB in general, synthetic biologists can choose a top-down or bottom-up approach to design living systems (Jewett and Forster, 2010). The top-down approach is based on taking an existing living system and trying to manipulate a limited range of functions. For instance, a gene can be inactivated - "gene knock-out" - to determine its role in an organism. This can help understand the whole network of genes this particular gene interacts with, through protein production, complex systems of activation and inhibition based on transcription factors and other mechanisms. This can for instance help in identifying the steps of metabolic pathways, these long cascades of chemical transformations which lead to the production of complex molecules inside cells. A knocked-out gene ceases to produce its associated protein, disrupting the metabolic pathway(s) which make(s) use of this protein.

On the other hand, synthetic biologists can also rely on the bottom-up approach. In order to build a new organism, this approach uses a *chassis* (B. L. Adams, 2016) - usually *Escherichia coli* or the simpler *Mycoplasma genitalium* - in which synthetic DNA will be inserted (Gibson et al., 2008). This allows for full control of the cell line being produced, but it is a more complex and error-prone process. It also allows for an iterative process of construction, making an organism more and more complex, as opposed to the iterative process of knocking out genes of the top-down approach. This technique was used in the creation of minimal-genome organisms (Hutchison, Chuang, et al., 2016; Hutchison, Peterson, et al., 1999). Minimal-genome organisms are precious research material as they contribute to the understanding of cell life cycles, whole-genome mechanisms, and the robustness of living systems.

Having presented what SB is, its principles, limitations, and applications, the current dissemination process of SB - teaching or popularizing - will be described in the next section. This will establish a reference to which Hero.Coli will be compared to.

## 2.5 Dissemination

The game Hero.Coli used later as a basis for this thesis was created as a new way to convey ideas from SB by taking inspiration from existing methods of SB dissemination and integrating them into a game. To give context to the process of design and creation of the game, the different ways by which the general public and students have been informed and instructed about SB will be briefly listed and analyzed.

### 2.5.1 Synthetic biology in popular culture

Controlling and modifying life was already a topic of interest for the general public even before the discovery of DNA and of its structure. The current most popular forms of culture through which the general public is exposed to SB, genetic engineering, and the ideas related to the manipulation of living systems are presented in this section. This will be a basis for an analysis on the level of information available to the public and on the types of portrayals of SB.

#### Science fiction literature

Mary Shelley's *Frankenstein* (1818) is both a founder of western science fiction's genre and an iconic example of fictional experimentation with life. H. G. Wells writings such as *The Island of Dr Moreau* (1896) bear the mark of his training in biology. One of his professors at what is now Imperial College London was Thomas Henry Huxley, a notorious proponent of Darwin's theory of evolution. *The Fatal Eggs* (1925) by Soviet physician and author Mikhail Bulgakov is a fiction depicting a scientist who discovers a ray which dramatically increases developmental growth rates in a wide range of species. It was published just a few years before DNA was hypothesized as the molecule carrying heredity through the works of Koltsov (1927) and Griffith (1928) (Aucamp et al., 2016). Interest in DNA spiked after the series of discoveries of the second half of the 20th century. Isaac Asimov, known for his *Robots* and *Foundation* cycles, wrote a few books about genetics, including *The Chemicals of Life* (1954), *The Wellsprings of Life* (1960), and *The Genetic Code* (1964).

The depiction of the manipulation of living systems, of the genetic mechanisms underlying life shows fear and fascination. The cliché of the unethical mad scientist, of the unfortunate creature freeing itself from its creator date back to the very first works of science fiction. But it is also worthy of note that the awe in front of the complexity and diversity born from Evolution are also a recurring theme.

#### Science-fiction films

Among the most successful recent popularizations of SB are the 1990 novel *Jurassic Park* by Michael Crichton - an influential sci-fi writer who has sold over 200 million books over his career - and its 1993 movie adaptation by Steven Spielberg which has grossed more than US\$1 billion worldwide. Neither of them is scientifically accurate. For instance, dinosaur DNA would be too degraded after 65 million years to be usable in a de-extinction process, and the scientific facilities shown are unrealistic: the laboratory shown in the movie has an unprotected egg incubator in the middle of an office. On the other hand, the fact that dinosaurs are shown featherless is consistent with the scientific consensus in 1993. But more importantly the essence of SB is reasonably well described. Some details are correct, like the fact that synthetic organisms can feature *kill switches* to enforce biocontainment. Recent de-extinction projects (Church and Regis, 2012; Maudet

et al., 2002) are reminiscent of the de-extinction of dinosaurs of *Jurassic Park*. The fact that geneticists in the movie fill in gaps of dinosaur DNA with frog DNA instead of bird DNA - birds are avian dinosaurs, they are closely related - is a plot device necessary for a reveal later on in the movie. Another example is *Gattaca* (Andrew Niccol, 1997). It portrays a dystopian society which practices eugenics and heavily relies on new technologies to control its citizens.

These movies have proven that the general public can be interested by stories featuring SB issues. None of them is naively optimistic about the potential uses of SB, quite the contrary: most of them end in disaster, underlining the need for control based on well-defined ethics. Unfortunately, blockbusters never really focus on short-term, realistic scenarios involving SB, and feature an highly unrealistic depiction of the work of SB engineers and scientists.

### Video games

Other media have also used the theme of SB, in fiction or in documentary formats. Many video games such as the *Deus Ex* series refer to genetics, gene therapy, genetic engineering, or genetic enhancement. Life simulation games partially overlaps with genetics, with games utilizing heredity, mutation, reproduction, and selection mechanics. Games like *Niche* (2016)<sup>16</sup> and *ARK: Survival Evolved* (2017) use simple game elements of *Tamagotchi* (1996)<sup>17</sup> while implementing selective breeding. Other games like *Gridworld* (2015)<sup>18</sup> or *Cell Lab: Evolution Sandbox* (2014)<sup>19</sup>, more simulations than games, feature all the traits of Darwinian evolution in a virtual ecosystem. Many other games just reference genetic engineering as a plot device, for instance to justify superpowers, or they reference it as a claim to be an educational game. On the Google Play store, at least 295 games have "evolution" and 94 have "genetics" in their title, a sizable part of which have no educational content. For instance, among the first results is *Alchemy ~Genetics* which is more of a simple animal hybridization game.

Again, in video games, most productions do not really focus on short-term, realistic applications of SB. But some games make it possible to play with genetic engineering and grasp an increasing number of notions and mechanisms as realism and complexity are increased in the implementation of the game.

### Public opinion

Concurrently to the increase of the depiction of SB in culture, the applications of SB have been met with resistance (Avellaneda and Hagen, 2016). A wide array of

<sup>16</sup> Steam page: [https://store.steampowered.com/app/440650/Niche\\_\\_a\\_genetics\\_survival\\_game/](https://store.steampowered.com/app/440650/Niche__a_genetics_survival_game/). "88% of the 1,615 user reviews for this game are positive" as of 2018-08.

<sup>17</sup> 76 million sold worldwide according to its Wikipedia page.

<sup>18</sup> Page on Steam: <https://store.steampowered.com/app/396890/Gridworld/>. "83% of the 183 user reviews for this game are positive" as of 2018-08.

<sup>19</sup> By Petter Sätterskog, updated in 2018. Main website: <http://www.cell-lab.net/>. More than a million installs, 50 thousand reviews on Google Play, rating it 3.7/5 [https://play.google.com/store/apps/details?id=com.satterskog.cell\\_lab](https://play.google.com/store/apps/details?id=com.satterskog.cell_lab).

issues have been debated for years, ranging from the least controversial - genetic testing to screen for hereditary diseases - to the most controversial ones - human cloning and eugenics.

The controversy over genetically modified organisms (GMOs), already described in part in section 2.4, culminated when a study intending on proving the carcinogenic nature of a variety of genetically modified corn was retracted (Séralini et al., 2014). This study focused on herbicide-resistant GM corn. This type of GMOs allows farmers to use a high amount of herbicides to destroy unwanted plants in the crops. This GM corn was given to mice and the group of mice was monitored. Several mice developed spectacular tumors. This study was retracted for a number of methodological reasons including the insufficient number of mice. Fundamentally, had the methodology been correct, Séralini and his team could only have proven that the food given to the mice caused the cancer to develop, not that it was necessarily the result of the genetic modification of the corn. But this variety of mice was already more susceptible to developing cancers, and the high amount of pesticides inside the corn given to the mice could also be a possible explanation of cancer development. GM plants rich in pesticides may be dangerous, like conventionally-bred plants rich in pesticides. Some GM plants may also be harmful in certain conditions due to horizontal gene transfer as discussed in section 2.4. But there is no study proving that GMOs *in general* are dangerous to health.

Another issue arose when one of the most prominent tools of gene editing, the CRISPR/Cas9 system, was demonstrated to be less accurate than previously thought (Lin et al., 2014). Stem cells research and gene therapy are other famous examples of contentious practices. It has to be noted though that the general audience is at least in part aware of the issues at stake, knowledgeable about the subject, and regarding positively the possible outcomes of SB. Besides, these opinions and levels of expertise vary over time and from a country to another. For instance, positive opinions and curiosity have been reported to increase in the Netherlands (Henneman et al., 2013). These variations in public opinion have resulted in varying degrees of legislation around SB research (Davison, 2010). The relations between public opinion, public knowledge of the field, and legislation on the field is still being debated (Marris, 2015), as well as plans and strategies to change the *status quo*.

### The need for SB dissemination

While the health-related, ethical, and ecological consequences of SB are being debated, its research and commercialization have carried on. The weight of SB in the US reaches "US\$350 billion per year, or roughly 2% of the US economy" (Church, Elowitz, et al., 2014), with a forecast growth of 14.5% on the 2017-2022 period<sup>20</sup>. This growth is often illustrated by the Carlson curve - the exponential

---

<sup>20</sup> "The global genome editing market is projected to reach USD 6.28 Billion by 2022 from USD 3.19 Billion in 2017, at a CAGR [compound annual growth rate] of 14.5% during the

reduction of cost in genome sequencing, see figure 2.10 - and compared to Moore's law (Ledford, 2010), illustrating the growth in digital technologies.

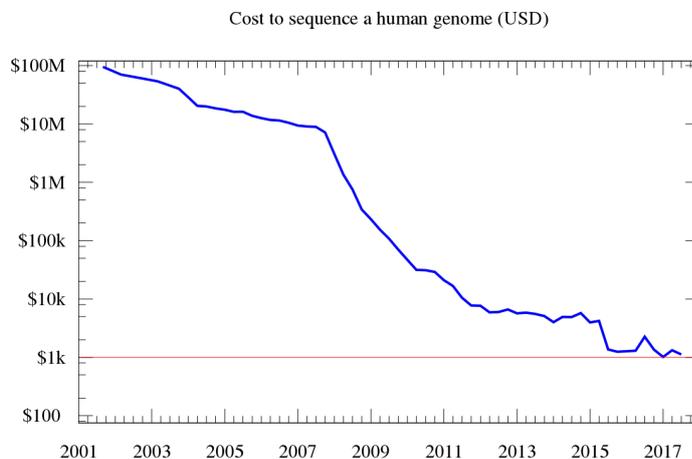


Figure 2.10: The Carlson Curve as of 2017

Source: Public Domain, <https://commons.wikimedia.org/w/index.php?curid=31006154> - Ben Moore, rewritten in gnuplot by grendel/khan.

As the SB sector is expecting a prolonged sustained growth, displays encouraging prospects in future technological developments, but raises concerns in the general public at the same time, it cannot be overlooked: awareness must be raised in the general public about the actual risks and prospects, to foster well-informed policy choices (Schmidt, Ganguli-Mitra, et al., 2009).

Interestingly, raising awareness is not as simple as it may sound. The way SB has been presented to the public from 2008 to 2013 has not yielded positive results (Pauwels, 2013). One explanation given by Marris (2015) is *SB-phobia-phobia*: the fact that by fearing that the general public distrust SB, scientists tend to present SB in a way that generates distrust in the general public.

The next section presents briefly how SB is taught to the new generation of synthetic biologists to demonstrate the range of methods used. This will be used later to show which ones can be integrated and how they can be integrated into Hero.Coli.

---

forecast period" (source: report "Genome Editing/Genome Engineering Market by Technology (CRISPR, TALEN, ZFN), Application (Cell Line Engineering, Animal Genetic Engineering, Plant Genetic Engineering), End User (Biotechnology & Pharmaceutical Companies, CROs) - Global Forecast 2022" published on <https://www.marketsandmarkets.com/Market-Reports/genome-editing-engineering-market-231037000.html>, last access 2018-08.)

### 2.5.2 Synthetic biology in academic training

This section is based on SB training as it is being practiced at the Center for Research and Interdisciplinarity (CRI) in Paris. Courses offered cover the Master 1 and Master 2 levels. Due to the similarities and the blurred boundaries between the subfields of biology described above, all of them are analyzed here. Their teaching indeed overlaps in part in terms of content and of methods of teaching. The different techniques of teaching are enumerated separately here, but in the training provided in the CRI, different techniques take place during a single lesson.

#### Lecture

Lectures are the most classical format for teaching: an educator will directly present the content they want the students to learn. It may be interspersed with questions to maintain engagement by the students and to follow on their progress and understanding of the content. Learning materials - academic papers, pictures, graphs, videos - are displayed and analyzed to illustrate and support learning.

#### Study of papers

Academic works can be assigned to students so that they learn about landmark historical papers and latest discoveries, about techniques and methods, and about how to present their work in the future when they write their own.

#### Wet lab experiments

Molecular biology experiments are conducted in a wet lab as a way to review the content learned previously, and to learn laboratory practices and techniques. This stage is a valuable training phase as they are at the core of the activity of synthetic biologists. This is also a way to learn more general scientific skills such as how to work in a team, how to follow a protocol, and how to conduct research. The required resources include a wide range of lab equipment, materials, sensors, and chemicals, shared with other teams and institutions using the same premises.

#### Design of genetic systems

As a practical and direct application, students may be required to learn about DNA sequence processing software utilized professionally by having a direct hands-on experience and designing genetic devices.

#### Modeling and simulation

Another key skill is implementing (wholly or partially) a model and running simulations. This activity is of course connected to the computational and systems biology aspects of SB. The programming languages used are Matlab or Python. The phenomena being modeled can be genetic devices such as the genetic toggle

switch (Gardner, Cantor, and Collins, 2000) or a genetic oscillator (Elowitz and Leibler, 2000). The models are usually based on ordinary differential equations (ODEs). They can help the students understand what are the methods, purposes, and limitations of biological models and simulations. The students can see how more comprehensive models are harder to implement and tweak due to the amount of parameters to tune. They can also see the differences between stochastic and deterministic models, and see in which cases one is preferable to the other. In the case of the toggle switch and oscillator, students can see that the simulations do not coincide with reality - in reality, genetic devices are less reliable due to the mechanisms which had to be set aside during the modeling phase - but they can give insights to the workings of the system, to its stability and to its steady-state(s).

### Games and simulations

A few games and gamified simulations have been published that cover SB topics. For instance, a simulation of the lactose operon has been developed (Esmaeili et al., 2015) and games such as Synmod (Schmidt, Radchuk, and Meinhart, 2014) can help students memorize the amino-acids and their properties. Online scientific crowdsourcing games like FoldIt (Cooper et al., 2010), EteRNA (J. Lee et al., 2014), Nanocrafter (Barone et al., 2015) and Phylo (Kawrykow et al., 2012) are used at the CRI to illustrate respectively protein folding, RNA folding, DNA folding, and DNA sequence alignment and phylogeny.

### iGEM

Students can also join the iGEM competition. This is outside of the formal academic training usually provided, but makes for a considerable experience in SB. Students committed to the competition will manage an engineering project, set objectives, meet deadlines, acquire various scientific meta-skills in addition to gaining SB skills. Finally, at the end of the project, the students may end up presenting their work in at the final iGEM competition meeting in Boston and meet and interact with their peers.

Having presented how SB is depicted in popular culture and how SB is taught in an academic setting, I will now present how SB is popularized.

### 2.5.3 Synthetic biology in popular science on the Internet

Some public institutions, private organizations and creators have also taken to themselves to inform the general public about SB, its known risks, capabilities, and prospects. This process can be motivated by multiple factors, such as the duty to inform the public about public-funded research, the will to raise awareness for educational or civic engagement purposes, self-promotion or personal interest in the field. Resources present on the Internet will be focused on as Hero.Coli was designed from the beginning as a primarily Internet-based free game. Their popu-

larity, their accessibility, and the quality level of those resources will be described, to later compare with the game we have realized at the CRI.

## MOOCs

The most formal level of online training that can be delivered to citizens is Massive Open Online Courses (MOOCs). They rely on the knowhow of professionals to be scripted, captured on video, edited, and published online. MOOCs typically also provide interactive support: learners' questions are answered in online forums, by other learners or by the team managing the MOOC. As an example, two collaborators at the CRI have implemented MOOCs pertaining to biology. Antoine Taly has created a MOOC called *Les origines moléculaires de la vie*<sup>21</sup> ("Molecular origins of life") while Edwin Wintermute has created *Synthetic Biology One*<sup>22</sup>. Wintermute's SB focused MOOC contains four parts, the first three being introductory courses to SB covering biology, maths, lab practice, human practice, DNA processing. The fourth part of Synthetic Biology One directly features Hero.Coli, the game used as backbone of the experimental part of this thesis. As of 2018-08, after a year of existence, 268 students have enrolled in the first course, 134 in the second, 70 in the third, and 292 in Hero.Coli; the YouTube channel Synthetic Biology One has 1.8K subscribers and 117K views.

## Videos

Videos on online video platforms bridge the gap between formal MOOCs offered by institutions and informal and often inaccurate formats for diffusion of knowledge in popular culture. For instance, dedicated online videos - including TV programs - can reach a large audience while providing it with basic understanding of SB mechanisms and issues. These videos can be hosted by YouTube or other video hosting services, and can be embedded in larger frameworks. For instance, students from the CRI, as part of the work for an iGEM project, have created videos hosted on YouTube that are accompanied by text that can be annotated. They grouped them into *MOOC IGEM High School* and specifically targeted highschool students wishing to inform them and give them an accessible way to learn about SB. The screenshot shown in figure 2.11 gives an example of one page from the MOOC IGEM High School. It consists of the text on the left with the clickable annotated expressions on a darker gray, the clicked annotated expression in yellow, its annotation in the center of the screen, and the video on the bottom right.

Annotations can consist of text, images and even videos. Anybody can annotate the text or monitor the activity on the page or on the MOOC. The Genius website that the authors used for this has originally been built to allow music fans to write and annotate lyrics, to document, give context and meaning to songs. MOOC IGEM High School creators have cleverly repurposed Genius and used it instead as a crowdsourced encyclopedia on SB with embedded videos.

---

<sup>21</sup> <https://www.fun-mooc.fr/courses/parisdiderot/56004/session01/about>

<sup>22</sup> <https://syntheticbiology1.com/>

**GENIUS** SIGN UP SIGN IN

FEATURED STORIES TOP SONGS VIDEOS COMMUNITY | SHOP | f | | @ |

**What are biobricks?**  
Igem Paris Bettencourt team

Featuring Registry of standard Biological Parts, BioBricks, Molecular biology & 5 more  
Produced by Igem Paris Bettencourt team  
Album Fundamental facts of Synthetic Biology

2:21

WHAT ARE BIOBRICKS? LYRICS

In this video, you will learn about the standardized biological parts known as Biobricks, how to use the biobrick database, and general advice on designing your constructs for your IGEN project.

Just as a reminder, BioBrick parts are standardized DNA sequences that can be used to design synthetic biology circuits in organisms like *E. coli*. Different biobrick parts include promoters, coding sequences, ribosomal binding sites, inverters, plasmid backbones, and terminator sequences.

We use them in synthetic biology because they are standardized and easy to assemble and reassemble, using simple molecular biology techniques. If you want to learn more about these parts, you can find more information on each of these parts in the supplemental resources provided or in a molecular biology textbook.

The BioBricks can be found in the open access Registry of Standard Biological Parts which was developed by researchers

Genius Annotation 1 contributor

What does it mean when it says "easy to assemble and reassemble"?

Before and after all BioBricks we can find standard sequences, the prefix and suffix sequences, which flank the 5' and 3' ends of the DNA part respectively. These standard sequences encode specific restriction enzyme sites. The prefix sequence encodes EcoRI (E) and XbaI (X) sites, while the suffix sequence encodes SpeI (S) and PstI (P) sites.

Genius Annotation 1 contributor

Standardized DNA Sequences  
PROMOTERS  
Coding sequences  
RBS

Figure 2.11: A screenshot of the MOOC IGEM High School page about BioBricks  
Source: <https://genius.com/Igem-paris-bettencourt-team-what-are-biobricks-annotated>

On YouTube, many professionals and amateurs have created videos about genetic engineering with various levels of polish and scientific accuracy. Some popular examples of such videos are shown in figure 2.12.

The three most viewed science popularization videos about genetic engineering were made by the science popularization channel *Kurzgesagt - In a Nutshell* (6.8 million subscribers). The first, *Genetic Engineering Will Change Everything Forever – CRISPR*, has more than 9.5 million views (all figures given as of 2018-08). *Genetic Engineering and Diseases – Gene Drive & Malaria* and *Are GMOs Good or Bad? Genetic Engineering & Our Food* have been watched respectively 5.3 and 4.1 million times. The most viewed dealing with awareness, ethics, and civic engagement is *Gene Editing: Last Week Tonight with John Oliver (HBO)* made by *Last Week Tonight*, with more than 4.5 million views.

### Other Internet resources

Other Internet resources available to citizens curious about SB include specialized website dealing with science popularization, ranging from amateur-managed sites to worldwide known academic publishing groups' such as *Science* (the American

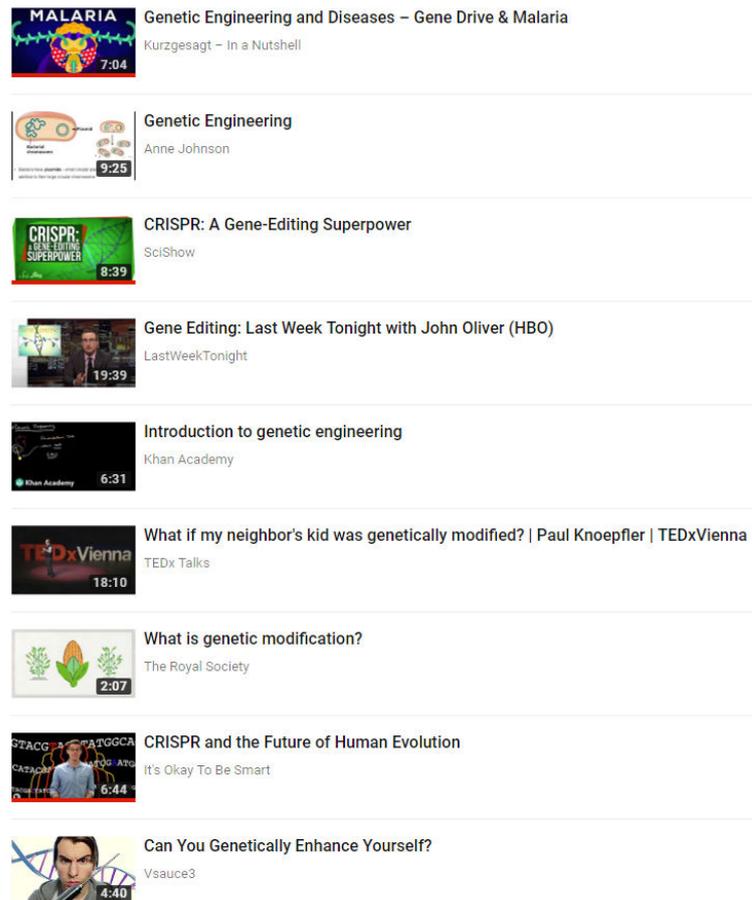


Figure 2.12: Examples of YouTube videos of SB popularization

Source: handpicked selection of videos returned by YouTube with the search words "genetic engineering" or "synthetic biology"

Association for the Advancement of Science group) and *Nature* (Springer Nature group). Dedicated wikis such as OpenWetWare<sup>23</sup> or the generalist wiki Wikipedia, provide easy navigation through hyperlinks, and extensive quality content due to their crowdsourcing process of creation. The main problems may be the technicality or low engagement of these formats. Online versions of newspapers and (science) magazines also deliver articles about SB from the technical or societal and ethical points of view to their huge audiences. Again, as stated in section 2.5.1, there is still a large proportion of citizens who are not aware of the existence of SB, and the process of dissemination itself still is not satisfying according to citizens themselves - trust still needs to be built between scientists and citizens (Pauwels, 2013).

Having listed the state-of-the-art characteristics of game-based learning, its tested approaches and techniques, as well as the content of the educational game

<sup>23</sup> <https://openwetware.org/>

Hero.Coli - synthetic biology - the contribution that this thesis intends on realizing will be stated in the next chapter.



# Chapter 3

## Questions, approaches, and objectives

### 3.1 Research questions

In a context of growing interest in synthetic biology (SB) and societal, technological, and economical consequences from SB research and applications described in chapter 2, this thesis proposes to demonstrate the complete process of design, implementation, and assessment of the SB educational video game Hero.Coli. In order to assess the impact of Hero.Coli, its effects will be explored and analyzed. This analysis, structured around the list of research questions below, was conducted in academic settings and in public experiments and is presented in the data gathering and data analysis chapters (chapters 5 and 6). This section details how these research questions will be evaluated: each of them will be answered using a list of hypotheses tested in the aforementioned chapters.

RQ 1: *Academic education*: What are game-based learning outcomes to SB education for University students in terms of knowledge acquisition and motivation?

RQ 2: *Popularization and lifelong learning*: What are game-based learning outcomes to SB popularization to citizens in terms of basic comprehension and interest?

These research questions can be evaluated by the following similar hypotheses:

H1a/*H2a*: the use of a game can increase the motivation and curiosity in learning, *resp. discovering*, synthetic biology.

H1b/*H2b*: the use of a game can make students, *resp. citizens*, understand basic notions about synthetic biology. These notions presented in Hero.Coli are:

- a) the simplified link between genotype and phenotype,

- b) the nature of BioBricks and devices: DNA sequences,
- c) the BioBrick simplified grammar (Promoter - RBS - CDS - Terminator),
- d) the names and functions of the bricks,
- e) the advanced notion of inducible promoters.

H1c/H2c: the use of a game can introduce misconceptions in the minds of students, *resp. citizens*.

RQ 3: *Learning efficiency, motivation and player characteristics*: Do players' characteristics - demographics, interests, practice - correlate with SB game-based learning efficiency in terms of knowledge acquisition and motivation?

H3a: age correlates negatively with learning outcomes; gender does not. The absence of correlation between learning outcomes and gender was shown in Papastergiou (2009) for instance.

H3b: age correlates negatively with motivation; gender does not.

H3c: learning outcomes correlate positively with interest and education in biology.

H3d: motivation correlates positively with interest and education in biology.

H3e: learning outcomes correlate positively with interest and practice in games.

H3f: motivation correlates positively with interest and practice in games.

RQ 4: *Playing duration and player characteristics*: Do players' characteristics - demographics, interests, practice - correlate with playing duration?

H4a: age correlates positively with playing duration to complete the game; gender does not.

H4b: playing duration to complete the game correlates negatively with interest and education in biology.

H4c: playing duration to complete the game correlates negatively with interest and practice in games.

RQ 5: *Player characteristics and implicit, explicit content*: How do the outcomes of different pedagogical strategies compare to one another?

H5a: explicitly-taught content is better understood and remembered than implicitly-taught content.

H5b: the young, students, gamers better understand and remember implicitly-taught content. Gender has no influence.

RQ 6: *Quiz-based assessment and automated tracking*: How comparable are learning metrics computed from questionnaires and from automated remote tracking data? Can quiz-based assessment be replaced by automated tracking?

H6a: In a linear game where each puzzle is compulsory to solve one after the other to finish the game, reaching some thresholds can be equivalent to validating the assimilation of a notion.

H6b: The learning outcomes are predictable using game metrics.

RQ 7: *Threshold effect*: Is there a threshold effect in the game, *i.e.* a point in the game after which no significant additional outcome is measured?

H7a: After reaching a certain point in the game, most of the learning is achieved: going further will not increase the learning outcomes.

## 3.2 Outlining the remaining chapters

Before evaluating these hypotheses, I will present in chapter 4 the design and implementation processes employed to create Hero.Coli. During these processes, the methods employed were borrowed from previous works on the field of educational games presented in chapter 1. Chapter 5 will present how the data gathering protocol was constructed and conducted in April 2018 in the Cité des Sciences museum, and in other previous instances. The data analysis process will be described in chapter 6. The conclusions drawn from this analysis will be summed up in chapter 7, alongside prospects of possible future developments and research leads to explore.



**Part II**

**Experimental Setup**



# Chapter 4

## Design and Implementation of Hero.Coli

This chapter presents how Hero.Coli was created in the Center for Research and Interdisciplinarity (CRI) in Paris. It was designed by CRI Master students during the academic year 2012-2013 as an entertaining way to share the knowledge and knowhow they had acquired in synthetic biology (SB). It was then implemented by CRI researchers, its goals being changed and adapted to the variable circumstances and new challenges that appeared throughout the project. The first concepts, principles, and designs are given alongside the successive improvements because most of the original ideas have been kept during the whole span of the project.

### 4.1 Genesis

The process of creation of Hero.Coli is deeply intertwined with that of the CRI. The first design choices and further developments were tied to research interests and opportunities. That is why it is sensible to briefly introduce Hero.Coli's context of creation. A complete timeline of the project is available in annex A.1.

#### 4.1.1 History of the CRI

The Center for Research and Interdisciplinarity (CRI), located in Paris, France, is affiliated with the Université Paris-Descartes (Paris 5) within one of Paris' associations of universities and higher education institutions, Sorbonne-Paris-Cité. It was founded in 2005 by Ariel Lindner and François Taddei, then researchers in the Inserm (French National Institute of Health and Medical Research), in the UMR1001 laboratory "Robustness and Evolvability of Life" (<http://www.u1001.org/>). Their work then focused on evolution, genetics, and the then nascent SB. By founding the CRI, they wanted to experiment with education, research methods, and new biotechnologies. In the recent years, the CRI has grown to now comprise more than 200 students in one license curriculum (bachelor's degree), two masters curricula, and two PhD programs. SB and education were the two

core disciplines that sparked the creation of the CRI - the third one, digital technologies, has been added later on. There are other transversal research themes such as Citizen Science (crowdsourced research projects) and the open source.

### 4.1.2 Synthetic biology

As described in chapter 2, section 2.5.2 and in the previous section, the CRI has had special interest in SB for years. SB is researched in the associated laboratory, shared internationally through MOOCs (see chapter 2, section 2.5.3) and through the annual participation of the Paris-Bettencourt team to the iGEM competition (see chapter 2, section 2.2.2), and taught in the two-year Master called AIV (*Approches Interdisciplinaires du Vivant*, "interdisciplinary approaches to the living systems"). Students are taught common scientific disciplines such as statistics, the reading and analysis of scientific articles, but also synthetic, systems, computational biology and laboratory methods and practice.

### 4.1.3 Education

In parallel with SB, the CRI research teams have also been interested in education since its creation. A few projects reflect this interest in education:

- the "EdTech" master level curriculum teaches education technology, pedagogy, in the light of 21st century skills;
- four "diplômes universitaires" (university degrees) are delivered by the CRI that pertain to education;
- the CRI project Savanturiers supports teachers in year-long projects to teach students from elementary school to high school the scientific method by practicing it;
- Sapiens, hosted at the CRI, is an inter-university service; its acronym means "Support Service for Innovative Pedagogies and Digital Education at Sorbonne Paris Cité [University]".

### 4.1.4 The Digital Synthetic Biology Club

In 2012, students of the AIV Master created a student club named "Digital Synthetic Biology", to work on the gamification of SB. These students, among whom Helena Shomar, started to write a draft of what would become Hero.Coli, in collaboration with CRI employees interested in video games or SB: game designer Benjamin Brogniart, 2d artist Armella Leung, director of research and co-director of the CRI Ariel Lindner, and synthetic biologist Edwin Wintermute. The development team understood the potential of an engaging SB-themed video game: as stated in chapter 2 section 2.5.1, SB was then nascent, promising, and lacking proper means of dissemination towards the general public. For instance there was

not a single video game available on the Internet about SB and BioBricks at that time. Even today, only a handful of games could claim to be SB video games. Part of the team had knowledge of SB and part of the team were video game players or experts, making it possible to share their knowledge in SB through an engaging video game. The game could also be an entertaining way of practicing SB and challenging colleagues inside the small community of synthetic biologists. The development team came to the idea of a crafting game based on genetic constructs with Role-Playing Game (RPG) aspects - relevant characteristics of RPGs are described in section 4.2.1. Crafting games such as Minecraft and engineering games such as Kerbal Space Program were starting to get extremely popular at that time, and because part of the activity of SB consists in engineering and crafting genetic systems using BioBricks (chapter 2, section 2.2.2). Novelty in the video game topic and the use of trending game mechanics would be the basis of Hero.Coli. An additional aspect that was present from the start of the project was that the CRI was also interested in education issues, and therefore Hero.Coli was also developed as an educational research project.

Hero.Coli as it was outlined was well inside the scope of topics of interest in the CRI. That is why the Hero.Coli development team took advantage of the participation of the CRI to the European project Citizen Cyberlab, which materialized the double specialization - SB and education - of the CRI. Hero.Coli's first development frame, Citizen Cyberlab, is presented in the next section.

### 4.1.5 Citizen Cyberlab

#### SynBio4All and RedWire/RedMetrics

From September 2012 to November 2015, the CRI was a partner of the EU-funded *Citizen Cyberlab* project (<https://cri-paris.org/citizen-cyberlab/>), a "European project focused primarily on exploring learning and creativity in citizen cyberscience" (Ramanauskaitė and Haklay, 2016), citizen cyberscience projects being "citizen science projects facilitated by the Internet" (Jennett, Kloetzer, et al., 2016). Citizen science in general comprises projects in which citizens contribute to scientific research through crowdsourcing. Each partner in Citizen Cyberlab had a delineated contribution associated with their competences and research interests. The list of Citizen Cyberlab projects can be found at <http://www.citizencyberlab.org/projects.html> (last access 2018-09). For instance, the CERN contributed by developing a crowdsourcing video game about the LHC called Virtual Atom Smasher (<http://test4theory.cern.ch/about/>).

The participation of the CRI consisted in two projects: SynBio4All and the RedWire/RedMetrics platform (Himmelstein, Couzic, et al., 2014; Himmelstein, Goujet, et al., 2016). RedWire is a user-friendly online game development platform for scientists, teachers and science popularizers, and RedMetrics is a game metrics and analytics tool briefly described in section 4.3.2. I was hired by the CRI in March 2013 to work on the RedWire platform. SynBio4All was a project encompassing CRI projects of SB popularization, among which Hero.Coli (referenced in

Jennett, Iacovides, et al. (2013) by its generic name "Synthetic Biology Game") and the MOOC iGEM High School mentioned in chapter 2, section 2.5.3. This explains why Hero.Coli was developed as a popularization tool rather than just a game using SB as a theme. As a popularization tool, Hero.Coli would need to address a large audience, who would need help in learning SB concepts in addition to learning how to play.

### Citizen Science

This thesis does not focus on Citizen Science, and SB crowdsourcing is not possible through Hero.Coli. However, the integration of crowdsourcing into Hero.Coli has been contemplated in addition to the objectives of popularization and education.

The game could allow the crowdsourcing of SB problems: given enough BioBricks and an accurate enough biochemical simulator, a set of problems could be solved by online volunteers, be them open-ended or not, already solved or not. For instance, contributors could try and create advanced genetic systems (oscillators, toggle switches, logic gates) with certain parameters. Following the example of FoldIt (Cooper et al., 2010), the protein-folding game that succeeded in implementing crowdsourcing using a video game, players would have to be trained first on well-known problems. FoldIt first provides its players with a step-by-step tutorial which teaches its game mechanics and controls, and some of the biochemistry involved. Most importantly, FoldIt explains to the player how they can win on a given challenge: by earning the best score, linked to the folding performed by the player. Similarly all of the challenges proposed in Hero.Coli would need an objective to be set, with a measurable outcome. A list of examples is given in table 4.1, from the simplest to the most complex to implement as a game developer.

The game would make it possible to explore the vast space of BioBrick combinations the same way FoldIt makes it possible to explore the vast space of protein folding combinations. It has been estimated that a single 150-amino-acid long protein can fold in  $10^{300}$  different ways (Levinthal, 1969). In the case of Hero.Coli, the combinatorial size of simplified BioBrick assembly - promoter, ribosome binding site (RBS), coding sequence (CDS), terminator - is roughly the product of the number of promoters, RBSs, CDSs and terminators, to the power of the number of genetic devices that can be injected into the bacterium. Since there are a few thousands BioBricks in total, this is clearly smaller, though difficult to explore and to test in various settings for different goals.

Bacteria generated by players would be evaluated through gameplay – either in FoldIt-like challenges, or in duels, or in online bacterium championships. The “best” bacteria according to a metric of the type defined in table 4.1 would be signaled to us, so that we could study them. A virtuous circle would ensue: either the real-life bacterium behaves *in vivo* the same way as it did *in silico*, in which case an interesting bacterium has been engineered; or the bacterium does not, meaning that the model that was used in the game has to be perfected, in which case the simulation can be improved, leading to an advance in bacterium simulation.

Objective	Type of problem	Success criterion
"make the bacterium produce protein $P$ "	closed problem: the player needs to use a genetic device containing the BioBrick coding for $P$	detect that $P$ is produced
"make the bacterium glow green"	closed problem: the player needs to use a genetic device containing the GFP BioBrick near blue light	detect the emission of green light
"make the bacterium blink green"	open-ended problem since there are several possible genetic oscillators	detect the intermittent emission of green light
"make the bacterium survive in this environment"	depending on the environment, may be any kind of problem	measure the longest virtual bacterium lifespan
"make the fastest bacterium"	open-ended and/or open problem; the challenge can be implemented either in a human vs computer or in a human vs human setting	measure the greatest instantaneous speed or shortest duration to reach a destination

Table 4.1: Examples of challenges that could be crowdsourced through Hero.Coli

## 4.2 Hero.Coli 1.12: a proof-of-concept

This section presents Hero.Coli in its first phase of existence, from September 2012 to November 2014. I joined the development team when the implementation started in June 2013. We developed a prototype (August 2013) which was expanded to a proof-of-concept (December 2013), followed by patched versions until November 2014. The development team comprised 8 members during the development of the prototype and 4 afterwards with a skill set including game design, programming, 2d and 3d art, sound design, and SB.

### 4.2.1 First design of Hero.Coli

#### Objectives

To summarize the objectives set forth in the previous sections, Hero.Coli was to be a video game:

- using the topic of **synthetic biology** (SB) and more specifically the use of **BioBricks**, due to the context of creation of the project;
- **popularizing** SB to fulfill Citizen Cyberlab's objectives;
- using **stealth learning** by offering an experience close to those of commercial off-the-shelf (COTS) games as it was hypothesized to be a better choice than *chocolate-covered broccoli* games;
- using **crafting** as a key game mechanic due to the obvious parallel with SB practices and due to the fact that crafting games were trending at that time;
- using **role-playing games** (RPGs) elements due to their widespread successful use in COTS games.

### Three Cs: Camera, Controls, Character

Since the first design of the game, the character has been a stylized *E. coli* bacterium. The camera is designed to be top-down, with an almost isometric view, centered on and following the avatar. The movement was first designed to be controlled using the right mouse button, while the keyboard was to be used only for shortcuts for menus and abilities.

The on-screen avatar and the camera were never changed throughout the project, but the controls were: during the first tests it appeared that several users were uncomfortable with the mouse. In the next versions, the users could configure the way they controlled the avatar - keyboard or mouse, using the left or right mouse button. No single solution was universally accepted, therefore after the proof-of-concept (version 1.12, November 2014) the keyboard and mouse were simultaneous possible movement controllers in order to maximize accessibility.

### Game genre

As the scientific content of the game was too complex for a short, casual game, and in order to encourage progression and foster play on a longer period of time, the game was designed as an adventure and exploration game with RPG elements. This means that the game had exploration features combined with challenges and puzzles which had to be solved to go on to the next zone. Moreover, puzzles based on SB would be an opportunity for players to demonstrate the skills they had acquired: a SB cell simulator gamified with puzzles unrelated to SB would risk being either a *chocolate-covered broccoli* or an entertaining, inefficient learning tool.

The RPG elements comprise the fact that the main characters follow a quest, that items have to be collected, that abilities are unlocked, and that the characters evolve and get stronger through an original fantasy world. In Hero.Coli, contrary to most RPGs, there is no experience system. The items that the player collects are DNA sequences. When equipped, these sequences modify the character itself,

instead of sequences just being a tool or a weapon. This kind of mechanics can be seen in some successful COTS video games, such as Deus Ex, BioShock, or Spore. What makes the specificity of Hero.Coli is that activating several of those items can yield unpredictable effects: usually, in games, items have direct, constant effects on the character's statistics. Some boots give a bonus in speed, armors give a bonus in defense. But in Hero.Coli, there is an intermediate layer of logic: the biochemical simulator described in section 4.2.2. This simulator takes into account the activated genetic devices, updates the chemical concentrations inside and outside the avatar, and finally computes the intensity of the abilities of the avatar. The avatar has a speed and a brightness that vary continuously between minimal and maximal values, instead of having discrete values like abilities do in most RPGs.

### Scenario and universe

The first design of the game has two main protagonists: Cellia, an *E. coli* bacterium, and an unnamed nanorobot. The nanorobot is the personification of the player in the game, visible only during the phases of narration - cutscenes, animations, pop-up windows. The nanorobot makes it possible to justify editing the genome of Cellia, akin to the metaphor of a sailor aboard a ship who can change the sails. Both of the protagonists are stranded in an unknown environment and have to advance throughout the level. This environment is a fantasy aquatic world with pearly white walls constraining the movement of the avatar, a blue background, drifting particles and unidentified creatures; static algae, mushrooms and flowers; red toxic clouds of the antibiotic ampicillin, flows of bubbles pushing the avatar, rocks that have to be moved, doors that open when shone upon with green light, and enemies moving on predefined paths that damage the avatar. This environment is not supposed to be realistic, its elements are either completely imaginary or unrealistic at the microscopic scale.

The objective of the game is stated explicitly at the beginning of the game: the introductory story given to the player describes Cellia and the nanorobot, the enemies, and talks about challenges that will have to be solved using synthetic biology, and the final goal announced is to beat an enemy - this is shown later on in figure 4.7, a screenshot of the game introduction. However, this proved to be a problem for casual players, who understood the final goal but frequently asked during demonstrations and playtests what they were supposed to do to play. The exploration and gathering of items was not intuitive enough.

### Game mechanics

The game uses game mechanics which are very common video game tropes: movement, gathering of items, exploration, "death zones" which kill the avatar, health and energy points, door unlocking, puzzle-solving by pushing objects to counter dangers - here, rocks block flows of bubbles which push the avatar into traps. The main originality of the game is the use of genetics and BioBricks.

In order to introduce the BioBricks to the player progressively, the game is structured on three levels of gameplay: (1) playing with single genetic devices (figure 4.1) (2) playing with BioBricks (figure 4.2), (3) playing with systems of interacting genetic devices (figure 4.3).

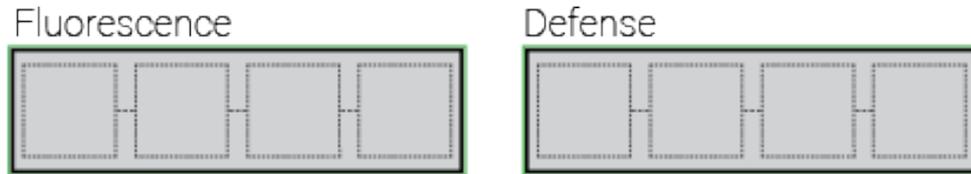


Figure 4.1: Genetic devices as abilities (first gameplay level)

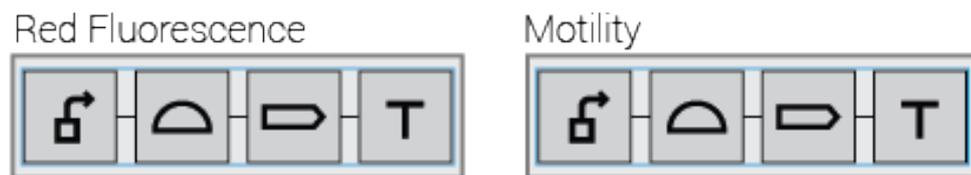


Figure 4.2: Genetic devices as independent BioBrick sequences (second gameplay level)



Figure 4.3: Genetic devices as interacting BioBrick sequences (third gameplay level)

In the context of Hero.Coli, contrary to SB, "genetic devices" have a precise definition: they are a construct of the type promoter - ribosome binding site (RBS) - coding sequence (CDS) - terminator, as shown in figure 4.2. More complex constructs are known in SB - the "RBS-CDS" subsequence can be repeated between the promoter and the terminator for instance. But in the game genetic devices can only produce one protein coded on the CDS, and be controlled by one promoter and one RBS.

When the CDS codes for a protein that directly has an effect, the genetic device directly controls an ability. For instance, the sequence constitutive promoter - RBS - Green Fluorescent Protein (GFP) CDS - double terminator presented in chapter 2 section 2.2.3 produces GFP when injected into a bacterium. In the

game, this sequence, when activated or *equiped* as typically described in RPGs, has the avatar produce GFP which makes the avatar glow green under blue light, by fluorescence. Other CDS BioBricks presented in table 4.2 produce proteins that have a direct action on the bacterium or its environment.

Coding Sequence	Protein	Function	Registry URL
Green Fluorescent Protein	GFP	Green fluorescence under blue light	<a href="http://parts.igem.org/Part:BBa_E0044">http://parts.igem.org/Part:BBa_E0044</a>
Red Fluorescent Protein	RFP	Red fluorescence under yellow light	<a href="http://parts.igem.org/Part:BBa_J06504">http://parts.igem.org/Part:BBa_J06504</a>
Flagella master regulator	FlhDC	Flagella synthesis, entailing increased motility <i>i.e.</i> movement speed: hexameric transcription factor activating flagella synthesis	<a href="http://parts.igem.org/Part:BBa_K343000">http://parts.igem.org/Part:BBa_K343000</a>
Ampicillin resistance cassette	$\beta$ -lactamase	Resistance to the ampicillin antibiotic: beta-lactamase (penicillin amidobeta-lactamhydrolase) catalyzes the hydrolysis of penicillins to penicilloic acids	<a href="http://parts.igem.org/Part:BBa_P1002">http://parts.igem.org/Part:BBa_P1002</a>

Table 4.2: Main coding sequences present in Hero.Coli

Therefore, in the first level of gameplay, players use genetic devices by equipping or un-equipping them, thus triggering the direct activation and deactivation of abilities. Figure 4.1 demonstrates the "black box" view of genetic devices, which are only considered in terms of abilities.

In the second level of gameplay, players use genetic devices at the BioBrick level: they replace BioBricks inside pre-existing constructs. They can increase the effect of a device by replacing the RBS BioBrick, which controls the level of expression of the genetic device. They can also replace the CDS BioBrick to switch abilities directly.

In the third level of gameplay, according to the planned game design, players were supposed to use genetic constructs that interact. This level of gameplay was not implemented. The idea was to show that advanced functions can be realized by equipping several devices at once. Interacting genetic devices can produce

complex behaviors, such as genetic oscillators and genetic toggle switches - genetic toggle switch devices have a constant high or low level of expression of proteins after a one-time stimulus, for instance the brief contact with a chemical substance. However, this level of gameplay, present at the very end of the level in a costly, carefully-scripted scenario, can only be mastered by players who have already mastered the two previous levels of gameplay. Many players were stuck in those two previous levels of gameplay during playtests and demonstrations, therefore the beginning of the level, overall game mechanics, and interface first needed to be fully grasped.

The third level of gameplay was never integrated in the main scenario of the game, but it has been integrated in another part. The ability of the simulator of Hero.Coli to correctly handle interacting devices and their emergent behaviors had been tested early in the development process, so they were integrated in the sandbox mode of a later version of Hero.Coli (see section 4.3.3). The sandbox mode is a common game mode: Kerbal Space Program and many other crafting and engineering games feature it. Minecraft even started as a *sandbox game* - there was no other game mode. Typically, the sandbox mode is a game mode in which most of the game elements are available to play with and in which it is not possible to win or lose, there are no clear rules to win or lose. Its purpose is to test out a game freely, without the constraints of a scenario or of a metric. In that regard, sandbox games and sandbox modes are *toys* rather than games: "it is the player who defines the rules which will generate entertainment" (E. Sanchez, 2014).

### Graphic User Interface and Crafting

In this version of the game, the avatar starts with one default ability which is not produced by a genetic device: movement. This makes it easier to start: there is no pre-existing device to explain to the player. The players have to pick up three DNA sequences: a motility device, to increase the movement speed from the low default speed to a medium speed; a GFP device, to be able to shine green in blue-lit areas; and an RBS BioBrick, in order to improve other genetic devices, in particular the motility device, to reach a high movement speed. Indeed, the use of these two devices and the crafting of a better movement device are compulsory to advance in the game: a puzzle requires medium speed to push a medium-sized obstacle, another requires high speed to push a huge obstacle, and a door-like element requires green light emitted using GFP to open. Therefore, players have to activate these two genetic devices and to craft at least once. Figure 4.4 demonstrates a typical situation in Hero.Coli where the player has equipped two devices visible on the left hand side in order to increase the movement speed and to glow green to open a door.

The first crafting interface is depicted in figure 4.5, with all interactive elements marked by a yellow numbered disk.

It is an extensive interface: BioBricks can be set on a *crafting table* using buttons (8) to (11) (BioBrick type selection) and (12), (13) (BioBrick selection), then crafted together (5) to create a genetic device (6), before *equipping* (activating)

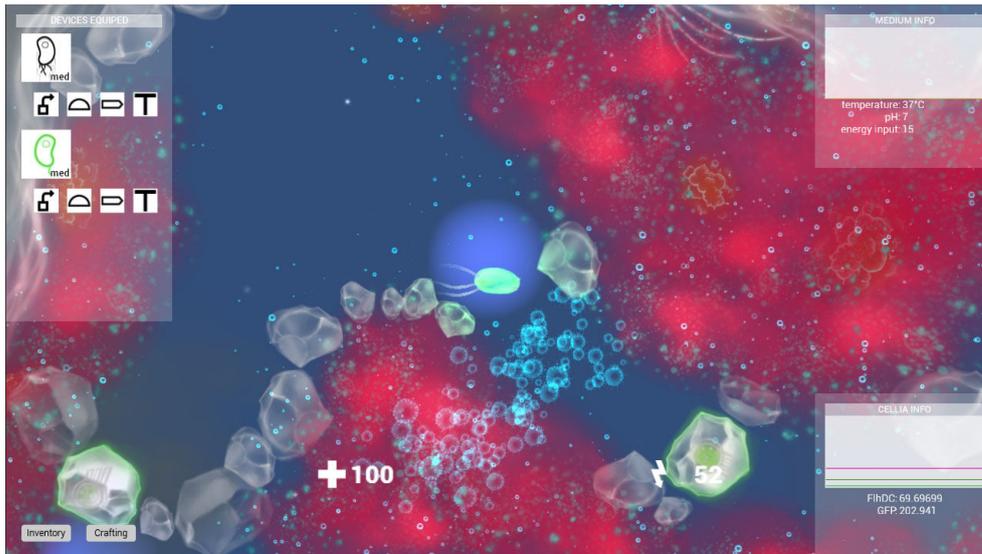


Figure 4.4: Hero.Coli 1.12: screenshot of an action phase

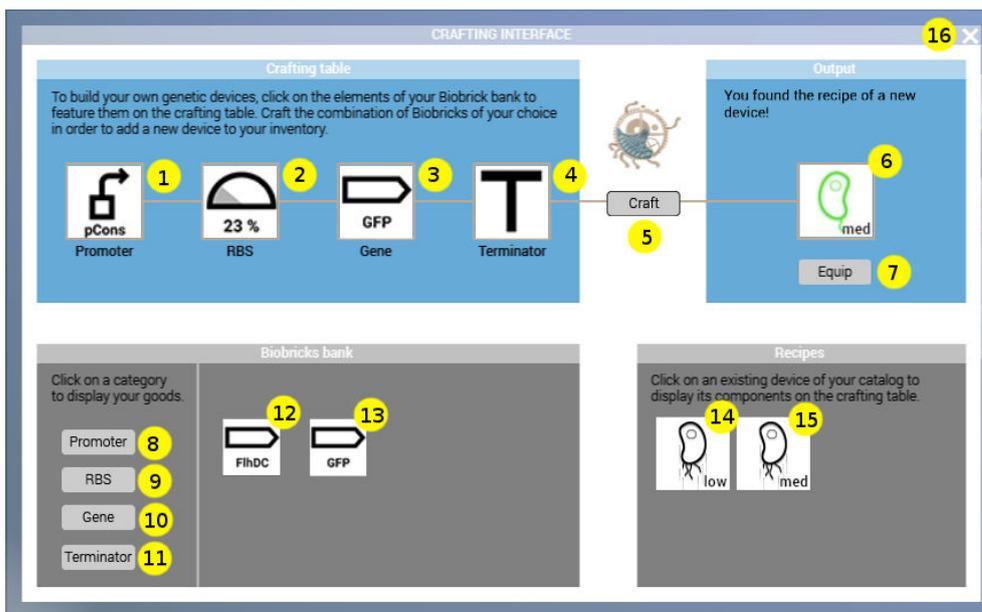


Figure 4.5: Hero.Coli 1.12: crafting interface

it (7) in a separate interface shown in figure 4.6. We chose not to use the drag and drop mechanic as it can be difficult for some users. Instead, the slot for the promoter BioBrick (1) is automatically filled in when the user clicks on a promoter in the BioBrick bank ((8) then (12),(13)), and so on with other BioBrick types. Crafting in that version does not limit either the number of available BioBricks once they are unlocked (picked up) by the player: when the player picks up the

new RBS, they can upgrade both the motility device and the green fluorescence device. There is no notion of stock of BioBricks, but rather of pattern or blueprint of BioBricks. There is also no limit on the number of genetic devices activated at any given time. The only limitation is that the same device can only be activated once. Otherwise, the player could reach arbitrarily high movement speeds and have the game crash. There is an additional limit that uses the mechanism of energy consumption: the more the active genetic devices are, the more the energy points decrease. The energy attribute is shown in figure 4.6, element (vi). When there is no energy left, the health points (v) decrease until the bacterium dies.

In the HUD (Head-Up Device, game interface visible during the action phases), figure 4.6, the crafting interface can be opened using the button labeled (D) in the bottom left corner.

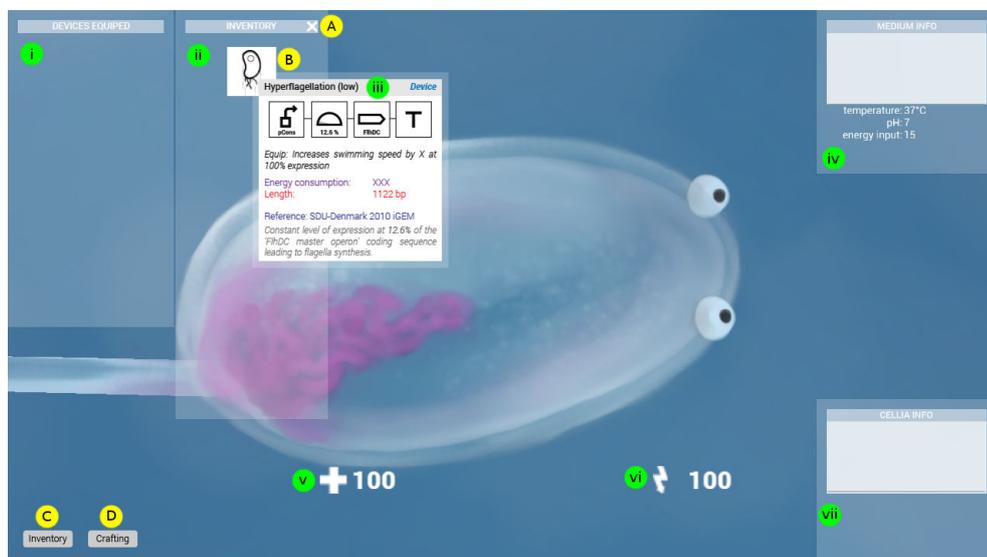


Figure 4.6: Hero.Coli 1.12: inventory and equipment interfaces, and HUD

Available genetic devices are listed in the inventory window (ii), which can be opened and closed using button (C), while already active devices are listed in the equipment window (i), which is always visible. The names *inventory* and *equipment* are conventional in RPGs. Button (A) closes the inventory, and the genetic device (B) is clickable: clicking it will inactivate the device. This screenshot also features a genetic device tooltip (iii) which appears when any genetic device is hovered with the mouse cursor. This tooltip gives additional information using icons and text, but players are not forced to read it. This is not the only part of the game where non-compulsory information is given. Windows (iv) and (vii), for instance, display the concentration of external and internal relevant chemicals respectively. This can help users identify what happens to the bacterium: a high concentration of ampicillin will tell them that the bacterium is being killed; presence of GFP inside the bacterium when it is not needed will remind them to deactivate the GFP genetic device.

During the playtests and demonstrations, we noticed that, even though the interface is thorough, it is not intuitive enough, as players either did not understand what they were supposed to do or did not know how to do what they wanted to do. We also noticed that the interface is burdensome, as many steps are necessary to craft and equip a device: starting from an empty *crafting table* (figure 4.5), without using the shortcut of clicking on an already-known pattern (14) or (15), a dozen clicks are necessary to craft and equip a new device. This problem was dealt with with new versions of the game, especially starting from version 1.50 upwards (see section 4.3).

## Tutorial

When the game starts, the first two pop-up windows introduce the scenario, global objective, and also the idea that the abilities of Celia can be edited by manipulating its DNA. The beginning of the game is designed as a minimal tutorial, to help the player discover the different aspects of the game, but only gameplay level by gameplay level.

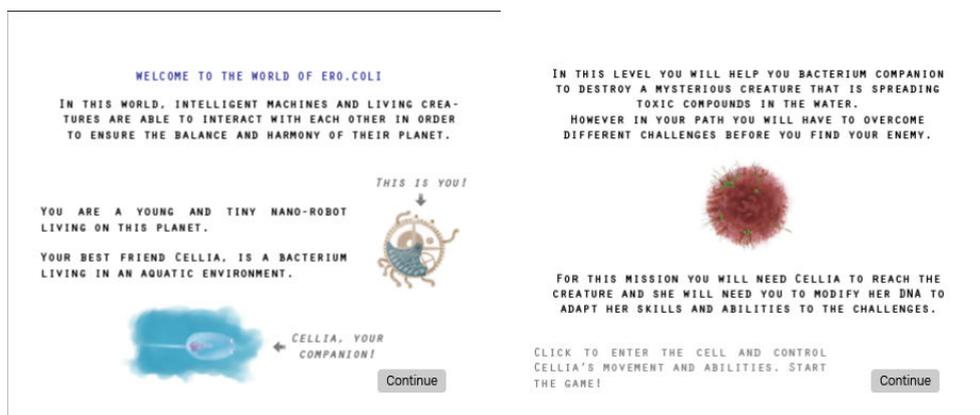


Figure 4.7: The first two messages displayed at the beginning of the game in the 1.12 version

**First step of understanding: genetic device and ability** First, in the first minute of play, the player picks up a genetic device. A pop-up is displayed which explicitly invites the player to “equip” that device by an on-screen message (figure 4.8).

When equipped, the device triggers the appearance of an additional flagellum, and increases the speed of the bacterium. The purpose of this part is to create an association in the mind of the player: equipping a device (motility device) means modifying the abilities (motility) and the appearance (flagellum) of the bacterium.

**Second level of understanding: BioBricks and crafting** Later on in the game, the player finds a BioBrick. This brick, an RBS, is able to change the power

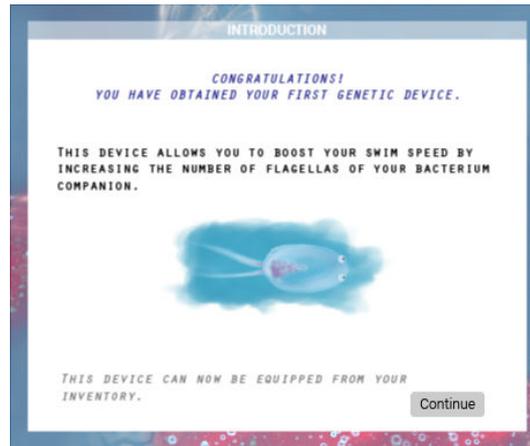


Figure 4.8: Device tutorial in the 1.12 version

of an ability when crafted into the device affecting the ability. The purpose of this part is naturally to make the player understand the role of this BioBrick, but also to show that the genetic devices present in this game are comprised of sub-components that control different aspects of an ability. This is how the concept of standardization (chapter 2, section 2.2.2) of BioBricks is taught implicitly to the player: by showing that BioBricks are interchangeable. This is also conveyed by the fact that BioBrick icons have the same shape. Similarly to the previous step, a message is displayed (figure 4.9).

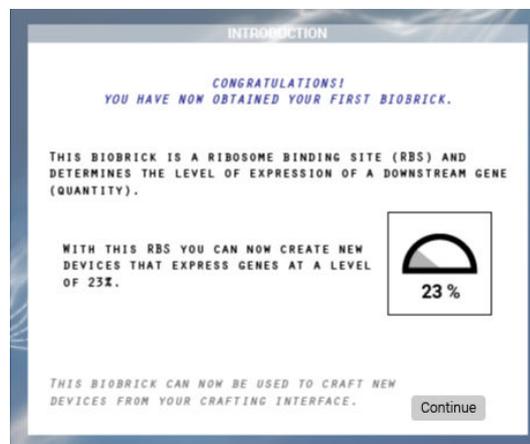


Figure 4.9: RBS BioBrick tutorial in the 1.12 version

Later on, the player finds another genetic device which, when equipped, makes it possible for the bacterium to glow green when illuminated by blue light. The player is supposed to find out by themselves how to use it, by reading tooltips (tooltips similar to the element (iii) shown in figure 4.6), by looking at the environment including a nearby blue light spot, by analogy with previous devices, by

comparing the BioBricks forming the motility and the fluorescence devices. The level design puts the player in a situation where they have to use the ability to go on.

### Game map

The game map was originally planned to be an open-world maze with high replay value to encourage players to explore again and again, to unlock BioBricks and game functionalities. For the prototype and then proof-of-concept of the game, we developed only a part of the planned map, shown in figure 4.10.

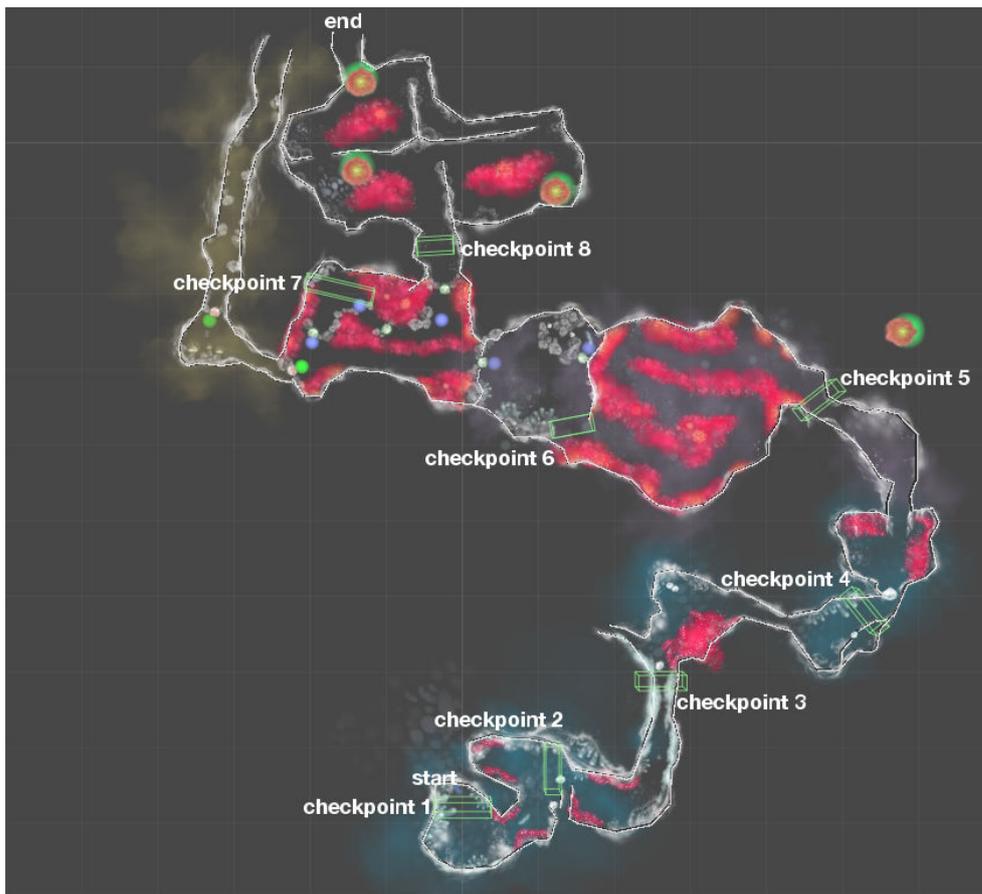


Figure 4.10: Map of Hero.Coli 1.12 with highlighted checkpoints

The path is rather linear, with the exception of a few dead-ends and open areas, a deadly trap, and the end area which offers two different paths. During the playtests and demonstrations, several players were confused by the end area and looped around, missing the exit. Others were confused or frustrated by the dead-ends. Only three passages are definitely locked when the avatar goes across them, making it possible to turn around and go the wrong way almost everywhere on the map. This proved also to be a problem, as it worsened the maze-like attributes

of the map. Gamers were accustomed to exploring mazes, but casual gamers were frustrated. The versions of Hero.Coli based on this map were improved little by little using elements that prevented players from getting lost or going the wrong way: one-way doors, rocks, bubble flows.

## 4.2.2 Simulator

What is referred to as the simulator here is the part of the code of Hero.Coli that computes the concentration of chemicals in different mediums of the game: inside the avatar, outside of it, and in some specific regions of the game map. The simulator used in Hero.Coli had been designed by Helena Shomar, member of the Digital Synthetic Biology club, with advice given by other members of the club, and by research director Ariel Lindner. A list of cellular biochemical mechanisms was drafted by them, with mathematical formulas quantifying each one. This simulator was then implemented, integrated and tested from March to August 2013.

Inside the avatar, each genetic system entails the production of proteins from its DNA sequence through the transcription – translation mechanism, the central dogma in biology described in chapter 2, section 2.1.1. After having computed each genetic system’s protein output, Hero.Coli’s simulator computes how those proteins interact with other proteins, and with other chemicals from the environment, and also takes into account some other mechanisms listed in chapter 2, section 2.2.3: protein expression, protein degradation, enzymatic activity, osmosis, active transport, and other chemical reactions, like for instance the hydrolysis of ampicillin by beta-lactamase. No other cellular mechanism was included in order to keep the gameplay simple. For instance, the division of the *E. coli* bacterium which happens regularly through the life cycle of the real bacterium would hinder the progression of the avatar or break the scenario altogether.

After the update of the concentration of all the chemicals, the abilities of the bacterium are computed from the final concentrations of the proteins. In this implementation, there is no solving of differential equations: the simulator, embedded into Unity, computes the variation of concentration on each time step, making the approximation, in the general case:

$$\frac{dP}{dt}(t) \approx \frac{\Delta P(t)}{\Delta t}$$

where  $P$  is the amount of a chemical and  $\Delta t$  the time step, typically between 10 and 50ms. In the particular case of protein degradation, by using its ordinary differential equation, the change in amount of  $P$  can be approximated to:

$$\frac{dP}{dt}(t) = -\lambda P(t) \approx \frac{\Delta P}{\Delta t}$$

$$\Delta P(t) \approx -\lambda P(t) * \Delta t$$

Unity indeed uses embedded loops of update of the different game subcomponents - the physics engine, the rendering engine, the UI,... - which are executed several times a second, depending on the complexity of the game and on the computational power of the machine on which the game runs. A simple game on a gamer PC can display a hundred frames a second, meaning that the rendering update loop executes every hundredth of a second, *i.e.* every 10 ms. Usually, frame rates do not drop below 20 frames per second, meaning a rendering time step of 50ms.

In order to configure and parameterize the reactions, chemical species, and mediums, the coder is required to fill in xml files. Reactions or phenomena that are not implemented into the simulator cannot be simulated directly, and must be approximated by other types of reactions.

### 4.2.3 Technical implementation

Hero.Coli was implemented using the Unity game engine with the C# programming language. We decided to make the code open-source to foster the creation of a community of Hero.Coli developers so that everybody could contribute by creating content for the game - BioBricks, challenges, or entire maps for instance. That is why the code is hosted on GitHub, a code-sharing online platform.

#### Open source

The CRI has developed several open-source projects, and promoted this practice across multiple student activities, among which the CRI fablab called the MakerLab, and the international synthetic biology competition iGEM mentioned in chapter 2, section 2.2.2.

That is why both the online platform RedWire and the games that are created using it are open-source. RedMetrics is a data-gathering tool, integrated within RedWire, which can also be used with any project as it offers an open API and open data. It is briefly described in section 4.3.2.

#### Game Engine: Unity

Game engines are the software foundation upon which game developers can create games. Game engines are the structure common to every video game: it usually but not systematically includes the management of sound, interactions with controllers - mouse, keyboard -, rendering, lighting, physics simulation... Without them, game developers would have to develop each game from scratch.

Unity was chosen early in the project to be the game engine, because of its numerous advantages. It is widely used in the video game industry, which tells a lot about its quality in terms of graphics, interaction, simulation, usage capabilities. More precisely:

- **A composite engine**

It combines several engines: a physics engine (a simulator), a lighting engine, a particle engine.

- **Component based design**

Unity is component based, meaning that the behavior of game objects can be described as bricks that can be added, removed, factored, shared amongst several game objects. It therefore allows for an iterative, step-by-step implementation, with a progressive complexification.

- **C#, high-level object-oriented language**

Unity can be used in pair with Microsoft's C#, a widespread high-level Object-Oriented Language (OOL), which adds to its ease of use, as OOLs tend to facilitate teamwork and robustness. It is worth noting that 3 members of our development team were proficient in C#-related languages like C++ and Java.

During the development of Hero.Coli, Unity upgraded its integrated development environment to allow a full integration of Microsoft's Visual Studio Code, a lightweight equivalent of the industry widespread Visual Studio.

- **An active engine**

Unity's development is active, with frequent major updates, as underlined in the previous paragraphs. Its other great subsequent developments during the span of the Hero.Coli project were a ready-made User Interface system, an improved lighting system, and even a Machine Learning agents module.

The most impactful development was on the game build output. At the beginning of the project, an online Unity game could only be played using the Unity plug-in. This browser plug-in required an authorization from the user when installed, by displaying a potentially alarming pop-up window in the browser. The typical reaction from non-engaged users was to leave the website without installing the plug-in, and thus without playing the game. This completely changed when major browsers such as Firefox and Chrome blocked the Unity plug-in due to security concerns raised by the Netscape Plug-in API<sup>1</sup>. Unity was forced to quickly develop and improve a new build process, making it possible to create plug-in-free games. This new process produces native JavaScript games, using HTML5's WebGL API that natively displays 2D and 3D graphics in the web browsers<sup>2</sup>.

- **An active community**

Unity has millions of users worldwide, including very active communities online and in real life. It makes it very easy to find answers to technical

---

<sup>1</sup> Browser point of view: <https://blog.chromium.org/2013/09/saying-goodbye-to-our-old-friend-npapi.html>. Unity point of view: <https://blogs.unity3d.com/2015/05/28/web-publishing-following-chrome-npapi-deprecation/>.

<sup>2</sup> WebGL described by its creators: <https://www.khronos.org/webgl/>.

questions or discover how to best make profit of Unity. Many websites such as Stackoverflow and the Unity forums offer detailed troubleshooting advice. Unity guides and Unity developer blogs provide beginners with step-by-step tutorials, completed with YouTube videos. Social networks pages advertise on new features and good practices. Finally, real-life clubs and Meetup events allow for even more awareness on Unity's capabilities.

- **A multi-target builder**

Unity makes it possible to easily build a game version for many platforms. The build available online is currently an HTML5 build. For demonstrations and experiments, Windows, Mac, Linux, and even Android builds have been used.

### Version Control System: Git

Version Control Systems (VCSs) make it possible to efficiently manage different versions of a project, whatever its content. In particular, different branches of development can be managed simultaneously - typically, feature branches and bugfix branches. A developer working alone on several aspects of a project can keep track of their advance, and different teams working on different aspects of a project can better coordinate. The VCS Git was chosen early in the development, partly because several members of the team were already accustomed to it, but also because it was already part of the practices at the CRI. This tool made it possible to work simultaneously on very different aspects of the game:

- assets: sounds, textures, animations
- bugfixing
- feature development

Our team comprised from 4 to 8 people for the development of the 1.12 version: blocking, urgent bugs could be fixed by some of us while the others could go on with other tasks, such as medium- or long-term feature development, or heavy refactoring that would otherwise disrupt each other's work.

VCSs are even more efficient for plain-text-based projects as they usually include tools to visually represent the differences between two versions. They highlight the edits between two successive versions or the discrepancies between two different branches of development, as shown in figure 4.11. In this case, on the left is the old version of the file and on the right the last, updated version. Removed characters and lines are represented in red, while the added ones are in green. The selected line is in yellow.

Unity's C# code itself is considered plain text and therefore can be very efficiently managed through Git diff tools. On the other hand, resources such as sounds, textures, animations, are considered *binary* and are thus not managed by diff tools. This was not a problem for us, as we made it so that only one person

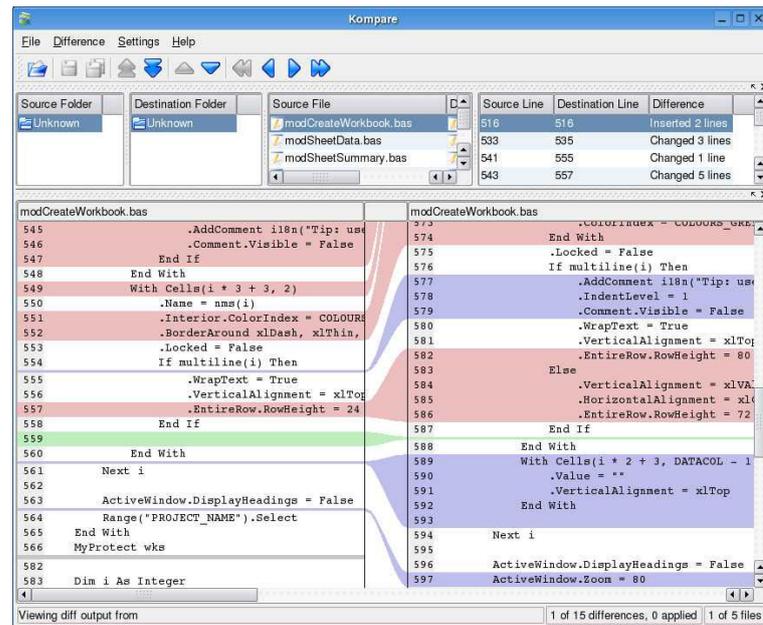


Figure 4.11: Screenshot of a *diff* tool

Source: 2004 screenshot of the diff tool Kompare <https://lukeplant.me.uk/blog/posts/kompare-a-tool-of-beauty/>

worked on one "binary" file at once - this never caused any major problem and did not cost us large development time. For a professional use with collaborative work on binary Unity files, there is a paying Unity CVS which makes it possible to visualize the differences between files considered binary by default.

We also used Git's online website, GitHub, to host our code and resources, and to organize our work<sup>3</sup> GitHub can be coupled with the ZenHub<sup>4</sup> plug-in to manage a project using the Agile / Scrum methodology. It makes it possible to list, label, assign, and process tasks. Moreover, ZenHub is free for open-source projects.

#### 4.2.4 Playtesting and accolades

Early on, we decided to apply part of the Agile methodology as mentioned in section 4.2.3. This methodology inherits from Rapid Application Development - also known as Rapid Prototyping in other fields. One of its mottoes is "test early and test often".

That is why we participated in several events early on and regularly along the development of the project.

<sup>3</sup> Hero.Coli game project on GitHub: <https://github.com/CyberCRI/Hero.Coli>. Data analysis project on GitHub: <https://github.com/CyberCRI/dataanalysis-herocoli-redmetrics>.

<sup>4</sup> ZenHub plug-in: <https://github.com/marketplace/zenhub>.

## Events

During the 2012-2014 phase of the project, we participated in:

- iGAM4ER 2013, Paris

After a few months of software development we participated in the International Game competition for Education and Research (iGAM4ER) competition in December 2013. This competition has been organized by the CRI almost every year since then<sup>5</sup> in the Cité des Sciences et de l'Industrie, Europe's biggest Science museum.

During this event, we could have guests play and comment on the game. The early feedback was positive: the game was awarded the design prize and the audience's choice prize. This early test with the general public made it possible to make adjustments to the game so that it better fit its intended audience.

It was also an opportunity to meet with creators of scientific games, who have gamer, game developer, and scientist point of views simultaneously, and therefore have more informed advice to share.

- FOSDEM 2014, Brussels

The Free and Open Source Developers' European Meeting (FOSDEM) is a yearly event in Brussels, Belgium, where software and hardware developers, officially nicknamed "hackers" share their experience on free and open source projects. We participated in this event, two months after the iGAM4ER competition, in order to meet independent game developers and open source programmers. We received positive feedback and interesting suggestions, showing that our project had appeal and potential among the tech-savvy.

- AMAZE 2014, Berlin

The AMAZE festival is a yearly video game festival that focuses on independent, creative, original video games. We intended on raising awareness about the game in the gamer community, and on getting feedback from the gamers.

- Cité des Sciences, 2014-09,10 Paris

The Cité des Sciences et de l'Industrie not only presents science-related exhibitions to the public and hosts events such as iGAM4ER as described above, but also hosts scientists who do research with the public. This scientist in residence program called Living Lab is organized by the Carrefour Numérique team of the museum<sup>6</sup>. We were hosted in September and October 2014<sup>7</sup> to

---

<sup>5</sup> Page about iGAM4ER on the CRI website: <https://cri-paris.org/igam4er/>.

<sup>6</sup> More on the Living Lab program, in French: <http://www.cite-sciences.fr/fr/au-programme/lieux-ressources/carrefour-numerique2/nous-connaitre/living-lab/>.

<sup>7</sup> Article by the Carrefour Numérique team: <http://carrefour-numerique.cite-sciences.fr/blog/herocoli-test-de-jeu-video-au-living-lab/>.

do a playtest with museum guests. This playtest is described in more detail in section 4.2.4.

- Science Museum LATES, 2014-05-28, London

Some of our collaborators of the Citizen Cyberlab project, working at the University College in London, presented Hero.Coli in this monthly, nocturnal event<sup>8</sup>. Hero.Coli was demonstrated on computers to the guests<sup>9</sup>

- GameCity 2014, Nottingham
- iGAM4ER 2014, Paris

### Playtesting results

Prior to the start of the PhD project, Hero.Coli was playtested at the iGAM4ER competition at the CRI in December 2013, and then at the Cité des Sciences et de l'Industrie in 2014. An online form was also published, in order to get additional feedback. This first gathering of feedbacks was intended to guide the development of the game, and not to gather scientific data. Therefore, the conclusions do not answer a scientific question and rather helped understand the potential for such a game, the flaws that had to be corrected, and the features that had to be developed.

The series of playtests we organized in the Cité des Sciences in September and October 2014 gathered 113 participants aged 6 to 62, with a mean age of 21.8 years old, comprising approximately 76.2% male participants and 23.8% female participants on the subsample of answering participants (self-assessed gender with three options: "male", "female" and "prefers not to answer"). The main conclusions and insights gained at the end of this first series of playtests were that:

- the game was appealing and engaging for a very diverse audience, from children to adults, mainly due to the gorgeous graphics; there was a strong divide between non-biologist gamers and non-gaming biologists as the former would often fail to understand, refuse to understand, or not care about science, while the latter would be unable to go through the first stages of the game which they deemed too difficult;
- many bugs made it difficult to go through and understand the game;
- in several situations, players would not understand what they were supposed to do, where they were supposed to go, or why such or such event happened. For instance players would usually not understand when the bacterium died of energy deprivation, causing them to feel frustrated. The solution we thought of was to expand the tutorial, add visual and audio feedbacks.

---

<sup>8</sup> On the Lates event in general: <https://www.sciencemuseum.org.uk/see-and-do/lates>  
On this particular instance of Lates: <http://mobilecollective.co.uk/events/>.

<sup>9</sup> YouTube video of this event: <https://youtu.be/FHKkjpP6mFI>. Hero.Coli can be seen at the 1:24 mark.

- many other elements - the controls, the interface, the content - could be improved or expanded using recommendations or remarks by the participants.

#### **Evolution of the project**

The game had then reached its final proof-of-concept stage. Players could combine BioBricks together, change the phenotype of the bacterium, and play from start to end. The game could be used in events to showcase our game development capabilities, to popularize SB, to have people come to us and discuss. At that point in time, Hero.Coli was already playable inside most Internet browsers – Firefox, Chrome, Internet Explorer, Safari. . . It was available online on herocoli.com and did not require an Internet connection when running. However, there was no offline executable file available, in order to encourage users to play while connected to the Internet.

Hero.Coli was to be re-purposed to be used as a basis for this thesis. The results from the fall 2014 playtests of section 4.2.4 were to help us decide which improvements to make in and around the game, from small features to a full rewriting of parts of the game and to the development of an assessment and analytics system.

## **4.3 Repurposing into a new research project: Adaptations**

From November 2014 on, Hero.Coli has been the focus of the PhD project presented in this thesis. Its aim, measuring different outcomes of playing Hero.Coli, required to push forward the development of the game in order to gather data on game-based learning on one hand, and to test different parameters on the other hand. We decided to keep the two objectives of popularization and education in order to reach out to different audiences: gamers, students, biologists, and the general public. The number of game developers involved has decreased to between one to three people at any given time since November 2014, however, the game was successfully adapted to fit the new design needed after the first playtests. The tracking system has been fully developed and embedded, the pretest-posttest has been facilitated online, the learning process has been made easier.

### **4.3.1 Accessibility**

#### **Platform**

As described in section 4.2.3, we have used Unity to develop the game. This made it easy to build versions for different platforms. We have decided to focus on the online version to gather data directly from the potentially huge online community. To make it possible, we have had to optimize the code and the assets - maps, sounds, textures. Offering a free online version of Hero.Coli playable directly from the browser also increases the number of potential players, since no installation is

necessary. Moreover, it forces players to be connected to play, making it possible for us to get tracking data but also to get answers to our online pre- and posttest surveys.

## Controls

As previous playtests at the Cité des Sciences had shown, lots of players wanted to be able to choose how to control the bacterium. When provided with only one way of controlling the bacterium, most of them complained; when provided with 3 parameters to control the bacterium on a test version – keyboard/mouse, relative/absolute orientation, and continuous/discrete clicks – the players chose uniformly in those categories. As a consequence, all modes of moving are now available in the game, and by default it is possible to use the mouse *and* the keyboard to move.

## Scenario

During the successive playtests and demonstrations, it appeared that the scenario was confusing. Taking into account the cost of adding phases of narration - cutscenes, animations, pop-up windows - we decided to just remove the nanorobot from the story. The nanorobot was kept only as the advisor who explains the game mechanics during the tutorials, and as an optional collectible item. These collectible items incite the player to play again until all of them have been found. It is a way for the player to show that they have mastered the game. All the other aspects of the scenario were kept: the fantasy world and all of its creatures, and the need to find the exit of the maze.

### 4.3.2 Metrics and analytics

At the beginning of the PhD project, RedMetrics was not developed yet. It was planned to become a data-gathering tool integrated within the RedWire online game development platform. RedWire's principles, architecture and functioning are described in Himmelstein, Goujet, et al. (2016).

## Data structure

In several entry points in the code of the game are calls to a function that sends data about an event - a position was reached, an item was picked up - to the RedMetrics server. These data describing an elementary event are called an observed element or *obsel* in the Trace Theory (Clauzel, Sehaba, and Prié, 2011). They contain the time at which an elementary event occurred, the location in the game - which chapter and at what coordinates, a code for the type of action the player performed, and complementary data, for instance the composition of the DNA strand the player acquired. These entry points are located directly before or after critical

events of the game: start, end, completion of level, acquisition of a DNA strand and so on.

In parallel to the metrics system, data format, and storage, I also had to make sure that users were properly identified uniquely, as this was not implemented in RedMetrics natively.

### **User Id system**

Developing some kind of user identification system was promising as it would make it possible to identify and characterize the behaviors of our users over a long period of use spanning several playing sessions, instead of just one. This was a high-priority problem: long-haul data gathering is critical to identify behaviors, therefore this identification system had to be implemented early on. We had the choice between three options: (1) by default, no identification system at all, as was then the case, with the previously underlined issues; (2) developing our own identification system; (3) using an already existing one. In order to spend the least amount of time developing and integrating this system into the already existing game ecosystem, we chose to consider only the Google sign-in system and a very simple home-developed identification system. We chose to develop the latter eventually.

### **Google ID**

The Google authentication system was considered before all other already existing such systems because of its reliability and widespread use in academic and general public contexts. The Facebook system was also considered, along with open source counterparts, but they were discarded because of their unbalanced use on academic and general public audiences. The Google sign-in would appear when users try to launch the game. Users would be invited to type in their Google email address and their password. Then, access to the game would be provided.

### **Advantages**

- Easy integration into our project's page. There are numerous online guides, tutorials, and examples on how to integrate it.
- Guaranteed unicity of the user across platforms and browsers. One Google user could play different versions of the game on different platforms – be it Android version on a phone or tablet, online version on Chrome, Firefox or any other browser, or standalone version – and still be identified as a unique user. This system would provide the best quality of information. Furthermore, a public space computer could be used by several users without them being tracked as the same user.
- Reliability as it is created and maintained by Google, and widely used.

## Drawbacks

- No control on the system at all. Google can change its API or its data policy at any time. As easy as it seems to integrate, this means that at any point in time, the whole system could stop working without any warning.
- Detrimental to the funnel process – the process that leads a user from the source mentioning the game – an article, a tweet, a QR code – to the actual gaming phase:

worst inconvenient: not all users do have a Google account. Those who don't probably won't take the time to create one either

it adds steps before being able to play the game: waiting for sign-in page, typing the email address, typing the password, signing in, waiting for confirmation

those steps may not be easily understandable: why a sign-in page? Which email address to enter?

some users may even fear that their personal data be stolen.

- Against open source and open data. We may have access to as much data as initially desired and designed, but providing Google with access to our data is nevertheless a threat. It is a bad practice in research to use these commercial solutions. Besides the fact that we don't have any control on the evolution of this system, we also have no clue how the system works internally. It is a black box making it impossible to know what happens to our data.

## Hero.Coli platform ID

The idea of this system is to assign a game-produced ID that is unique for a specific device – and browser when applicable. For instance, there is one unique such platform ID for every browser installed on a computer, and one unique platform ID per Android device, and one for every standalone game installed. The platform ID is created when the user first connects to the Hero.Coli website, stored locally, and sent using RedMetrics with the *start* game event. That way, the platform ID and the session ID are associated and stored.

## Advantages

- Transparency. Users don't have to sign in, the funnel process is not broken. It's so transparent that currently there's the problem that the users don't even know about the tracking system, which is a problem on an open-source, open data project. However, these data are completely anonymous, and it is planned that we add a message about this on the webpage of the game.

- Complete control on the system. We can make it evolve the way we want and adapt it to our very needs. We can simplify it or make it more complex; we can make it lightweight and quick, or comprehensive and slow.
- Per-platform tracking. This was not a primitive goal for this system, but one of its positive side effects is that it makes it possible to classify cohorts of users according to the platform they use.

### Drawbacks

- Maintenance. Has to be maintained by our team if the external APIs we use are modified, or if a bug is found.
- Incomplete information. There is no way to identify a unique user across platforms, or different users on the same platform. However, it appeared that public space computers – such as those in high schools and universities – use sessions. These sessions allow those platform IDs to differ from one user to another on the same computer. So, for the use we are considering, this drawback is not too crippling. The most important factors – development time and accessibility – led to choosing the second option, the platform ID, as most of this system had already been developed. The tradeoff implied letting reliability and comprehensive information aside.

#### 4.3.3 Academic use

In order to be used in class or as part of a curriculum, Hero.Coli had to be adapted. First, its content had to be greatly enriched in terms of the number of available BioBricks. Then, the context in which BioBricks could be used had to be modified. A calmer and more controlled environment had to be set in the game. In the former version of the game, the path of the bacterium was dangerous and stressful. This prevented the game from being used regularly in the classroom because of the excitement it could cause, and also because of the detrimental effect on concentration, even more so when some players felt stressed out because they were being watched while playing. Consequently, it was decided that a sandbox mode would be developed, in which all of the BioBricks of the scenario mode would be available right away, and in which there would be almost no danger, only some interactive elements such as rocks, doors, and diverse chemical clouds. Such sandbox modes are very common in commercial video games, and notably in engineering or crafting games like *MineCraft*, *Kerbal Space Program* and *Besiege* in which there are plenty of bricks to test out. In future versions of Hero.Coli, this free-roaming space to experiment with the game could be complemented by a series of independent levels which could focus on specific SB concepts or constructs, by using for instance themes described in MOOCs.

The sandbox mode helps create a learning ecosystem in the CRI: students study in the classroom, experiment in the wet labs, learn at home on MOOCs, and

experiment in Hero.Coli. Non-students can also follow MOOCs and experiment in Hero.Coli, or play Hero.Coli and then follow MOOCs out of curiosity: this way, MOOCs and Hero.Coli make each other benefit from their own audience.

### 4.3.4 Game design

#### Level design

Figure 4.12) demonstrates the new game map, its checkpoints, chapters, and path, including loops and U-turns that the players must accomplish to complete the game. The level was redone in order to allow for a more progressive introduction of game elements as described later in the tutorials section 4.3.5. This new design does not feature dead-ends and traps which frustrated players in version 1.12 (section 4.2.1). There are also many more elements that prevent the players from going the wrong way as often as possible. Frequent checkpoints are strategically placed to prevent the players from having a frustrating experience of losing their progress when their avatar dies. Loops have been introduced to induce a feeling a familiarity when the player goes back to an already-explored section to open a previously closed door. This was hypothesized to give casual gamers self-assurance, while still using game tropes familiar to other players. Additionally, as the parts of the level act as hubs or loops, they are re-used, the player has to go across those locations at least twice. Thus, using loops was also a way to make the game longer while saving development cost by reducing the level-building time. Level building is time consuming because it includes the assembly of 3D elements, textures, lights, the scripting of interactive elements, and debugging.

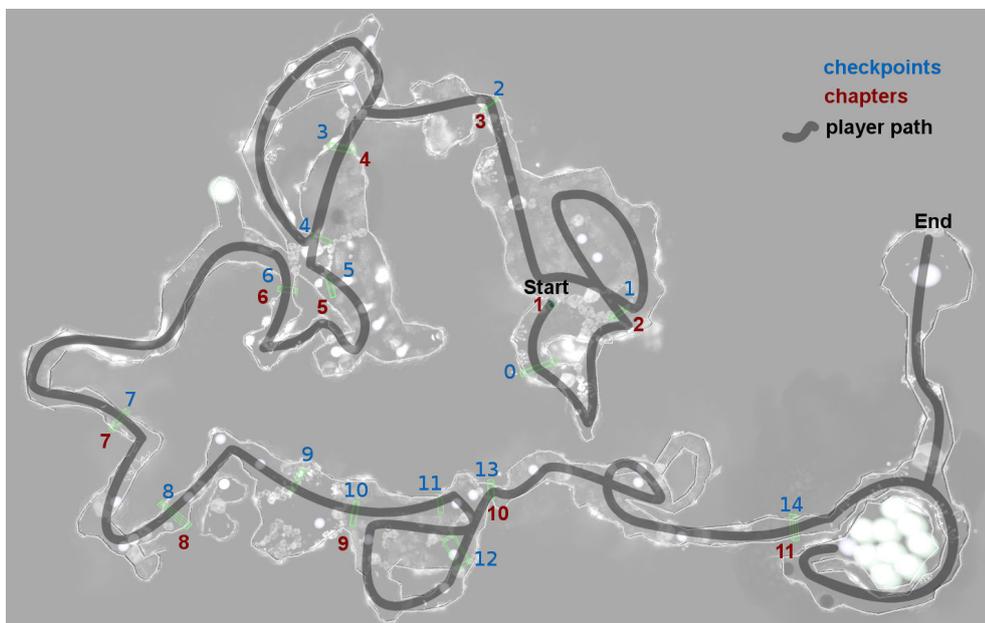


Figure 4.12: Map of the game with highlighted path, checkpoints, and chapters

### Increased use of crafting

We decided to change the way BioBricks were obtained and managed during the game.

In the first iteration of the game, BioBricks were *patterns*, that is to say, once a brick was picked up, the cell could produce as many copies of the brick as needed, as mentioned in section 4.2.1. For instance, if the player acquired one RBS, they could change all of their genetic devices by replacing their old RBSs with the new one. The problem was that this reduced the number of times the player picked up bricks, and also reduced the control we had as game designers over the enhancement of the cell.

In this new iteration, we made bricks countable: a newly-acquired RBS could only be used in one genetic construct at once. This made it possible to (1) have the players think strategically about their stock of BioBricks (which BioBrick is missing? which device should be prioritized and received the most powerful RBS?) (2) tightly control the rhythm of enhancement of the bacterium, and make it as progressive as necessary. Consequence (2) is also a way to respect the cognitive overload avoidance principle (Sweller, 1999; Kirschner, 2002) widely used in pedagogical applications and UX design of video games, which states that the players will not learn if they receive too much information over a short time period. This principle was more generally used in tutorials, briefly described in the next section.

### 4.3.5 Tutorial

The popularization objective of the game was to be fulfilled by developing even more the tutorial itself. This way, non-biologists would not feel lost or overwhelmed by the scientific aspects while the non-gamers would not feel frustrated by the difficulty of the game.

#### Step-by-step granular tutorials

To further avoid cognitive overload as described in section 4.3.4, Hero.Coli's text-based tutorials have been replaced by simple tutorials where the player is prompted to perform defined actions, with clear visual elements highlighting relevant parts of the screen and no more than two sentences displayed on screen.

The tutorials are structured in *steps* that can be skipped but with a cost: the player has to move the cursor on different zones on the screen. Players who are not engaged in the tutorial may skip the steps by clicking the highlighted game elements without thinking, but the text of the tutorial will be displayed, the game elements will be highlighted, and the players will have clicked on the elements we wanted them to click on. We hypothesize that this mechanic is better than a text that can be skipped: the player will at least see part of the game elements and know which elements can be clicked on.

In order to further avoid cognitive overload, we strove to do *granular* tutorials, *i.e.* tutorials that demonstrate the fewest aspects of the game at once. Only game

elements that cannot be dissociated are presented together to prevent the players to associate two independent elements. For instance, new BioBricks are presented one by one, because giving a new device comprising several new BioBricks at once could have the player think that these BioBricks have to work together or are somewhat related.

Finally, we also added cutscenes to demonstrate some game mechanics in addition to a pop-up with text and an image, in order to address all types of audiences: the players who do not want to read and those who do. This is an application of the well-known principle "show, don't tell": demonstrating instead of describing.

### Adaptive tutorials

Two blocking points in the game, *i.e.* parts of the game where players were stuck, now feature adaptive tutorials akin to the adaptive feedbacks studied in Conati and Manske (2009) with very simple triggers : if the player has not accomplished an action in a defined time interval after an event has occurred, a window appears which explains what must be done to get unstuck. The two instances are: (1) at the very beginning of the game, in chapter 1, when the player is supposed to start moving, (2) in chapter 2, when the player has just replaced their movement ability by their fluorescence ability - some players did not know what to do to be able to move again. In the first case, the player is prompted to use the keyboard or the mouse after 10 seconds of immobility. In the second case, the player is told how to activate the move ability back after 15 seconds without activation of the ability.

## 4.3.6 Interface

### Inventory, Equipment, Crafting

The whole interface was simplified. Redundant windows have been removed or merged into other windows: the inventory has been merged into the crafting interface and the equipment has been removed - see figure 4.13. The crafting interface shown in figure 4.14 has been simplified in order to reduce the number of clicks necessary to create a device and equip it. Depending on the precise state of the crafting interface, the crafting and equipping process now needs around four clicks - one per brick -, instead of a dozen (see section 4.2.1). In figure 4.13, the player needs to click on two available BioBricks on the bottom left corner - one RBS and one terminator - to complete the highlighted incomplete device being assembled on the top left corner. When the fourth BioBrick is put in its slot, an animation plays that joins the four BioBricks and results in a genetic device as seen above the highlighted incomplete device. If the device has never been assembled before, a new device icon appears on the bottom right corner.

This process of reducing the number of necessary clicks was a successful application of the UX principle of "minimizing [the] physical load" (Hodent, 2017). Additionally, applying another principle of UX which is the use of *signs* (ibid.), we replaced text by icons whenever possible, for instance on device icons in fig-

### 4.3. REPURPOSING INTO A NEW RESEARCH PROJECT: ADAPTATIONS<sup>107</sup>

ure 4.13. Previously, in version 1.12 (see figure 4.5) the level of expression was shown using "low" or "med". In version 1.50, the level of expression is shown using yellow chevrons.

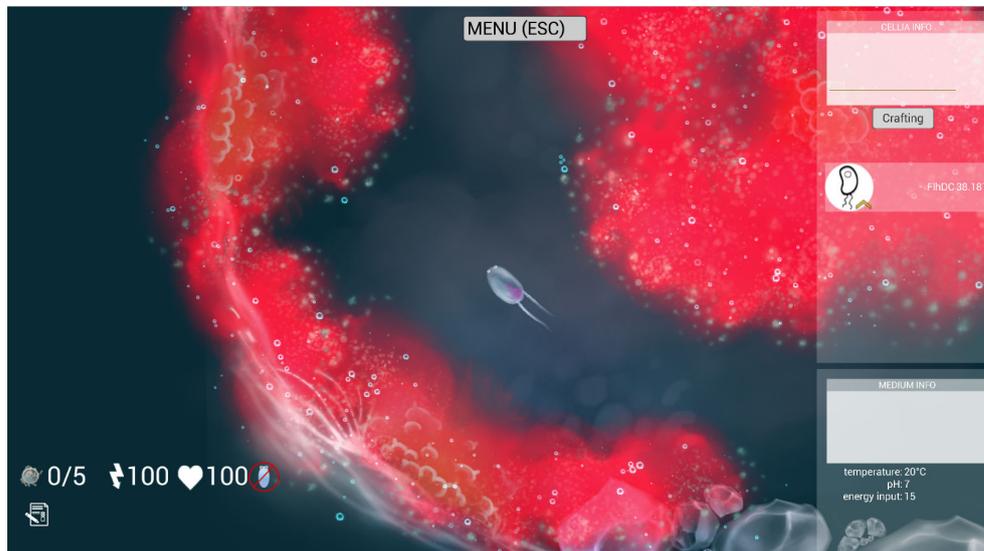


Figure 4.13: Hero.Coli 1.50: screenshot of an action phase

#### 4.3.7 Simulator

As an exploratory exercise, I have implemented a whole cell simulator with the help of computational biologist Vincent Danos. This simulator implements the model from Weiße et al. (2015) depicted in chapter 2, figure 2.9. It includes “cell housekeeping” (damage repair and renewal, waste management), cell division, and the energy metabolic pathway. I experimented by integrating it into test versions of Hero.Coli, but I chose not to integrate it in the online version of the game as this would have been detrimental to the gameplay. Cell division had already been set aside when designing the first biochemical simulator as described in section 4.2.2. The energy metabolic pathway and cell housekeeping mechanisms were too difficult to integrate in the current game scenario because they would introduce more game mechanisms, making failure cases difficult to identify. For instance, an insufficient energy output could stall the movement of the bacterium, and a failing housekeeping system could prevent the cell machinery to function and produce proteins: it would be difficult to differentiate a problem caused by an inadequate genetic device from one caused by a failing cell. More generally these mechanisms would make the game teach cell biology notions to the players while its objective was to teach SB notions, making the game confusing. However, this cell simulator can still be integrated in a future version of Hero.Coli or in a separate game centered on cell biology.

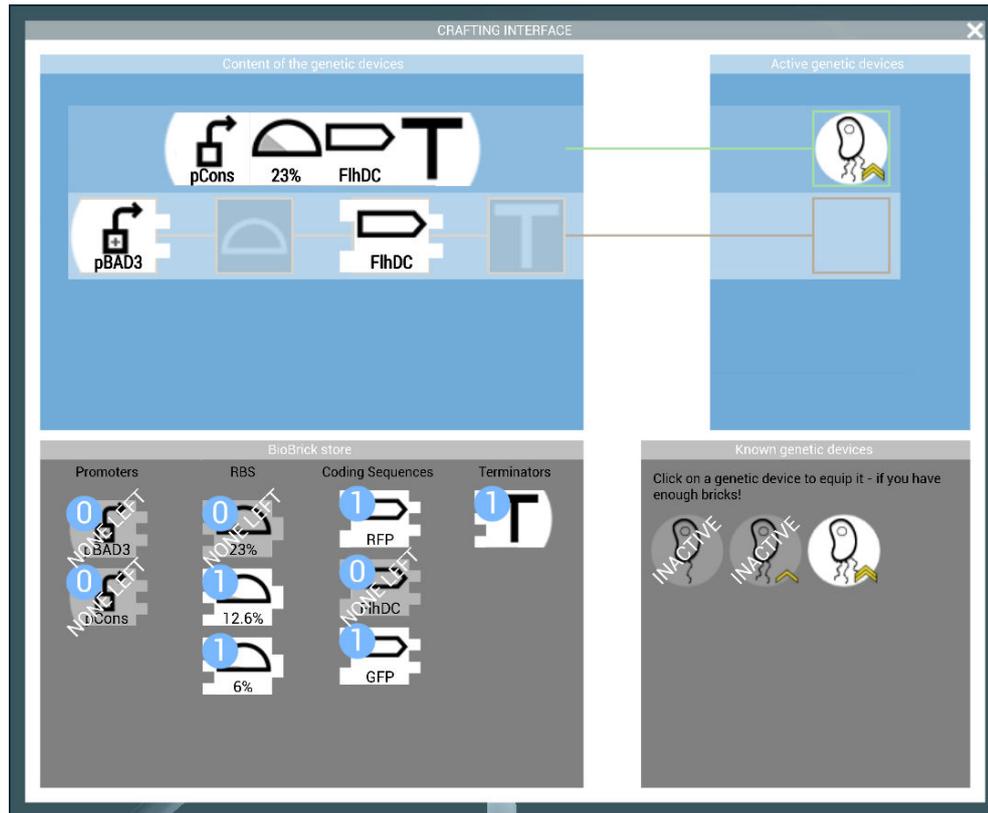


Figure 4.14: Hero.Coli 1.50: crafting interface

Besides this experiment, the biochemical simulator has not been dramatically changed. It was improved by fixing its bugs, by implementing the last planned features, and by making it easier to tweak.

Having seen the creation process, characteristics, and development environment of the SB video game Hero.Coli, we will see in the next chapter the different iterations in the data gathering process and in the elaboration of the experimental protocol. Following the principles of Design-Based Research, Hero.Coli, its tracking system, and assessment system have evolved together, therefore the evolution of the game seen in this chapter is mirrored by the evolution of the assessment process in the next.

## Part III

# Data gathering and analysis



# Chapter 5

## Data gathering campaigns

The methods implemented to gather data are presented in this chapter. These methods will not describe in detail the tracking system RedMetrics (Himmelstein, Goujet, et al., 2016) used to track Hero.Coli game events. Rather it will focus on the surveys produced in pretest-posttest experiments, communication efforts, and the online, in-class, and public experimentations. The final section, section 5.4, presents the final experiment conducted in the Cité des Sciences museum in April 2018. Its results are presented in the next chapter, chapter 6.

### 5.1 Survey methodology

Hero.Coli's effect was assessed using a pretest-posttest design (Dimitrov and Rumrill, 2003) through online pretest and posttest surveys in all experimental contexts presented in this chapter - online, at school, or in public events. Participants filled in the pretest survey, played the game for at least half an hour, and then filled in the posttest survey (which was the same as the pretest survey in most of the implementations). There is a priming effect when using this methodology: participants will be aware of the theme of the game, parts of its vocabulary, mechanics, and graphical elements. This gives time to participants to get accustomed to the game universe, and to know in advance what to look for in the game. The priming effect was not tested in any experiment as part of this thesis.

The questions were designed in order to get usable information about the participants' background and about their learning and enjoyment of the game. All questions are self-assessed, for three reasons: (1) in order to allow the experimenters to manage larger groups: 2 experimenters can have a group of 5 people start participating at once; (2) in order not to influence the answers of the test subjects, especially concerning the age, gender, and interest questions; (3) in order to have experimental conditions close to those of online players, as Hero.Coli is an online game. The questions were designed to cover most of the content of Hero.Coli while keeping the list of questions as short as possible.

Indeed, the questions were also designed to keep the participants engaged. This ensures that their answers will be usable and meaningful: from the very

first preparatory experiments presented in this chapter, discouraged, unengaged participants were observed to suddenly give up reading the questions and quickly answer, thus not demonstrating their understanding of the game. Questions of increasing difficulty were assumed to help keep the participants engaged in the assessment process. For instance, the very last question of the surveys of the final experiment was put at this position in the surveys because it was expected to be discouragingly difficult:

Q: Last question. Next page only contains remarks. Guess: you have crafted a functional device containing an arabinose-induced promoter and an arabinose Coding Sequence (CDS). What will happen?

1. It produces arabinose all the time
2. It is active only in arabinose clouds
3. It produces more and more arabinose after being induced, because it induces itself
4. It produces nothing since it induces itself
5. I don't know

It is the longest question and it pertains to a concept that had already been identified as difficult to grasp by players - promoter induction.

To keep the online and offline participants engaged, we also decided to make part of the pretest survey optional: we had noticed during the first experiments that museum experiments worked better when a critical mass of participants was present in the room. Especially for groups of people who want to do the same activities together but are not equally motivated: in order to convince the group to participate, some members are allowed not to fill the entire survey. We also applied this logic to online participants because online participants are not as engaged as museum visitors.

The usability of the participants' answers was increased for the final experiment due to an improvement in the presentation of the options. Indeed, in earlier surveys, the listing of possible answers for each question was randomized: the "I don't know" answer could be located anywhere in the five vertically-displayed possible answers (see figure 5.1). The reason behind this randomized positioning is that I wanted to get rid of the position bias (Blunch, 1984). Unengaged participants show apathy, expend less energy, and would consequently minimize the time spent moving the cursor and selecting an answer in every group of possible answers by picking random, inconsistent and therefore unusable answers. In the final experiment, the appearance of the questions was changed to prevent these hard-to-spot random answers. The "I don't know" option is now always proposed as the last option, on the right side of the questionnaire - see figure 5.2. This way, an unengaged participant may easily select all the "I don't know" options on the same horizontal position of the screen while scrolling through the questions. It incidentally makes it easier for the biology-uninformed too. Generally, it incites participants not to answer randomly while making it easier for them to click "I don't know".

What are BioBricks and devices? \*

- DNA sequences
- I don't know
- Proteins
- Amino-acids
- RNA sequences

Figure 5.1: Survey: randomized vertical position for possible answers

	RNA sequences	Amino-acids	Proteins	DNA sequences	I don't know
What are BioBricks and devices?	<input type="radio"/>				

Figure 5.2: Survey: constant horizontal position for possible answers

The questions of the final experiment listed in annex C.4 are used to assess the participants' learning and follow the hypotheses set forth in chapter 3:

H1a/H2a: "the use of a game can increase the motivation and curiosity in learning, *resp. discovering*, synthetic biology": questions 1 and 5.

H1b/H2b: "the use of a game can make students, *resp. citizens*, understand basic notions about synthetic biology", these notions being:

- a) "the simplified link between genotype and phenotype": questions 11, 17;
- b) "the nature of BioBricks and devices: DNA sequences": questions 12;
- c) "the BioBrick simplified grammar (Promoter - RBS - CDS - Terminator)": the 8 sub-questions of question 15;
- d) "the names and functions of the bricks": the 13 sub-questions of question 14;
- e) "the advanced notion of inducible promoters": sub-questions "device 3" and "device 4" of question 15, question 18.

The questions assessing the motivation of the participants are questions 1, 9 (indirectly), 10. Questions 2, 3, 4, 5, 6, 7, 8 are used to determine the profile of the participants: age, gender, and interest and practice in video games and biology.

The remaining questions assess whether the rest of the content of the game was learned: questions 13, 16.

## 5.2 Online experiments

In order to prepare the pipeline of data gathering and analysis for well-controlled experiments, we conducted preliminary online experiments. From the very first events registered on RedMetrics, the analytics process has indeed been progressively refined by making use of a constant flow of online participants, generated through regular advertisement, demonstration and exhibition. Hero.Coli and its associated research project as a whole were continuously exhibited during the entire repurposing and ulterior development. The aim was to detect flaws in the experimental protocol: flaws in the game design, in the implementation, in the data gathering or data analysis. Following the principles of Design-Based Research (Wang and Hannafin, 2005; Amiel and Reeves, 2008), the game, the experimental protocol, and the data gathering and analysis pipelines were constantly improved by taking into account participant feedback and survey and tracking data. Meeting varied audiences was critical to make sure that all paths of the experimental pipeline were tested out. Here is a list of events during which Hero.Coli was presented, and their audience:

- AMAZE 2015, Berlin: gamers.
- Epic Genetics Day 2015, San Jose, USA: museum guests.
- OuiShare 2015, Paris: tech enthusiasts, innovators.
- Stunfest 2015, Rennes, France: gamers.
- OpenFiesta 2015, Shenzhen, China: students, teachers, researchers.
- RUE 2015, Paris: university professionals.
- FOSDEM 2016, Brussels: developers.
- FOSDEM 2017, Brussels: developers.
- L'Échappée Volée, Paris: tech enthusiasts, innovators.
- Futur en Seine 2017, Paris: general public.

Additionally, we used the CRI website, social networks (Twitter <https://twitter.com/herocoli> and Facebook <https://www.facebook.com/herocoli.game>), the popular link-sharing website Reddit, and the citizen science platform SciStarter (<https://scistarter.com/project/18858-HeroColi>).

The development of the system means that an important part of the data of the early online experiments was unusable due to technical adjustments. Moreover, only a fraction of the online participants answered both the pretest and the posttest, of which only a fraction answered the optional part of the pretest survey as said in section 5.1. Concretely, during the 2017-06 to 2018-03-22 period, the pretest and posttest data of only 46 out of 474 participants were usable. This

sample was not big enough to ascertain definite conclusions on the related research questions through the conclusive use of data analysis tools (used later in chapter 6, section 6.2.3), but it made it possible to prepare the final experiment in the Cité des Sciences presented in section 5.4. Similarly-sized groups of such online participants were surveyed in later online experiments, with the same ratio of exploitable data.

This low rate of exploitable data may be explained, in addition to the concurrent development of the tracking system, by the fact that filling surveys is a tedious task not expected by online participants attracted by the appeal of an eye-catching free online video game about genetics. Indeed, contrary to the participants to real-life experiments, online participants are not encouraged by peer pressure or constrained by a teacher to do a task, they participate as long as they wish. Therefore the online participants usually do not fill in the posttest survey. In future pretest-posttest online experiments, extrinsic rewards could be used to have the participants fill in the posttest survey. Incentives would be implemented, such as for instance unlockable content accessible only to those who filled in the posttest survey (new game map, new character, new music), an online high score table, or a list of achievements and badges underlining the contributions and involvement of players.

### 5.3 In-class experiments

To see how students reacted to the game, how to have them make the most of the experience of playing Hero.Coli, we conducted the following series of in-class playtests and preparatory experiments: on 2015-12-14 (14 students aged 14-15), 2015-12-16 (17 students aged 14-15), 2016-02-01 (10 students aged 16-18), 2016-04-26 (27 students aged 18-22), 2016-05-04 (32 students aged 16-18), 2016-09-08 (10 students aged 14-15), 2016-09-15 (11 students aged 14-16), 2017-10-16 (25 students aged 17-19), 2017-11-21 (17 students aged 21-22). For instance, we noticed that students tended to watch each other play, exchange tips with or without understanding the game mechanics, and collaborate to answer the pretest and posttest surveys. The issues they had while playing were taken into account to prepare the next versions of the game. These in-class experimentations helped me prepare the protocol for the final experiment in the Cité des Sciences presented in section 5.4.

Some of these early experiments demonstrated that the students learned the content of the game, but also learned a few misconceptions. In the questionnaire shown in annex C.1 that accompanied Hero.Coli 1.12, questions 16, 24, 26 targeted possible misconceptions introduced by the game: respectively, the fact that the avatar, a stylized *E. coli* bacterium, has eyes; the fact that the avatar had to push rocks and swim against bubble currents; the fact that the avatar and DNA strands were of comparable width and height. Figures 5.3 and 5.4 show the results of question 16 "E. coli has eyes" in pretest and posttest on the sample of 32 students aged 16-18 who played on 2016-05-04 and were surveyed immediately before and after playing. The correct answer is "no", in light green. However, the design of the

question itself can explain the dramatic change of answer from a correct majority in pretest to an incorrect majority in posttest. The name "E. coli" in the question can be interpreted as the real *E. coli* or the game avatar *E. coli*, and many students may have missed the earlier mention in the questionnaire "the following questions refer to real-life E. coli". That mistake was corrected in the next version in 2016-06 (see annex C.2) which systematically mentions "real E. coli".

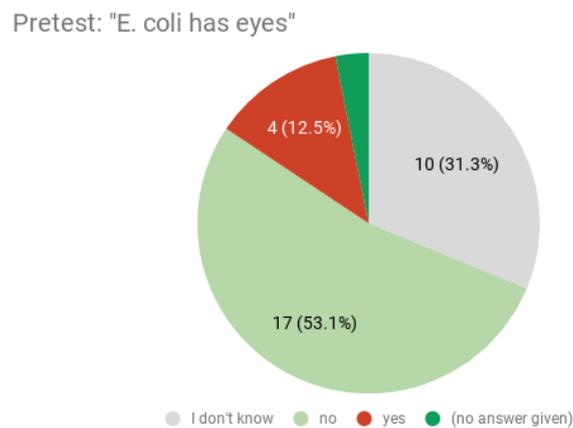


Figure 5.3: Results of question 16 - pretest

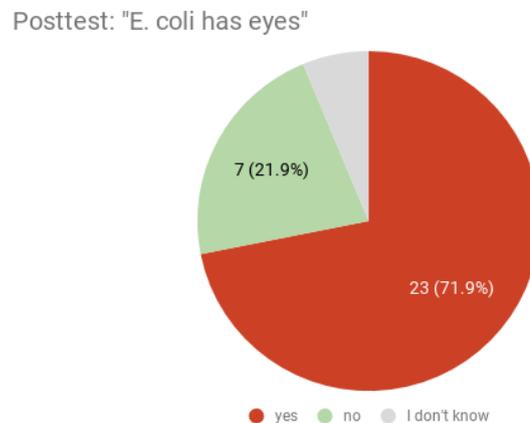


Figure 5.4: Results of question 16 - posttest

Figure 5.5 and 5.6 demonstrate an example of successful learning. The question is question 12, "*Which biobrick controls what is produced by the genetic device?*". The correct answer is "*ORF*", in light green. These results must be put into perspective by the fact that there was no "I don't know" option. Students could choose not to answer as one student did in the pretest, but they usually chose

one, possibly because they did not realize that not answering was an option. Also note that at this point, the Coding Sequence (CDS, see section 2.2.2) was referred to as the ORF, the Open Reading Frame, which, however a related notion, is erroneous and was corrected in the next version. This significant increase in score can be explained by the studious nature of the experiment - at school, with the biology teacher attending the experiment - and by the fact that such groups of students usually did share answers to one another. The priming effect discussed in section 5.1 may also explain part of the increase.

Pretest: "Which biobrick controls what is produced by the genetic device?"

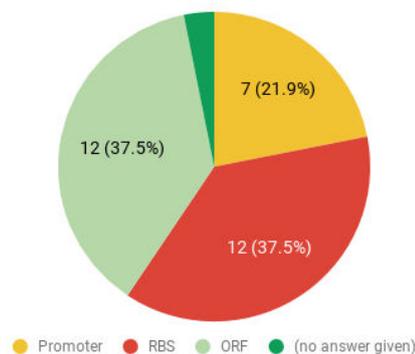


Figure 5.5: Hero.Coli 1.12: pretest answers to "Which biobrick controls what is produced by the genetic device?"

Posttest: "Which biobrick controls what is produced by the genetic device?"

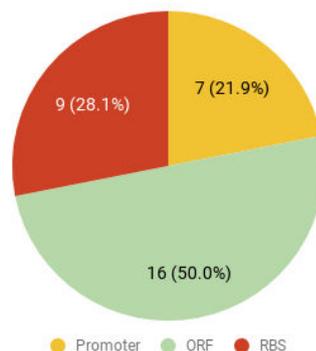


Figure 5.6: Hero.Coli 1.12: posttest answers to "Which biobrick controls what is produced by the genetic device?"

During this preparatory experiment we also gathered preliminary evidence to answer research question RQ5: "How do the outcomes of different pedagogical

strategies compare to one another?" which tended to lead to the conclusion that there was no significant difference between implicit and explicit learning strategies in version 1.12 of the game. In this version, some game elements are explicitly taught through compulsory pop-up tutorials while other are not: RBS BioBricks are explicitly taught while CDS BioBricks and the fluorescence mechanism are not. These implicit elements can either be inferred by the participants, or known through a tooltip that appears when hovering a CDS or a GFP CDS respectively. By comparing the results from explicitly- and implicitly-taught game elements which are comparable we can have an idea of the difference of impact. Figure 5.7 and 5.8 are the pretest and posttest answers to question 11: "*Which biobrick controls only the efficiency - level of expression - of the genetic device?*". The correct answer is "*RBS*" in light green. The pretest and posttest values are comparable between question 11 (explicit learning) and question 12 (implicit learning). This was speculated to be due to many factors: (1) a negligible impact of this format of tutorials (which we therefore edited), (2) an undetected difference in nature between the two concepts between tested (CDS and RBS) maybe due to the in-game integration - one concept may be more frequently-used than the other for instance, (3) a problem in the implementation of the protocol. A better protocol to test this research question would be to have two versions of the game: one with explicit tutorials and one without, and to compare the outcomes on the same questions in both versions.

Pretest: "Which biobrick controls only the efficiency - level of expression - of the genetic device?"

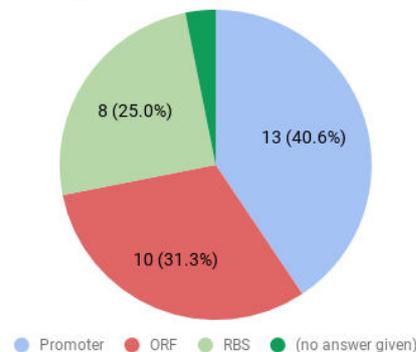


Figure 5.7: Hero.Coli 1.12: pretest answers to "Which biobrick controls only the efficiency - level of expression - of the genetic device?"

**Gathering results before April 2018** The online and in-class early experiments enabled us to establish a protocol and guide our tracking and game development efforts. They however resulted in insufficient and incomplete data that could not be exploited to conclusively show all of the effects induced by the game,

Posttest: "Which biobrick controls only the efficiency - level of expression - of the genetic device?"

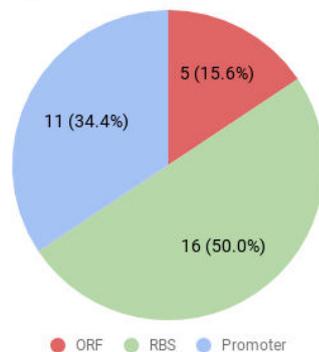


Figure 5.8: Hero.Coli 1.12: posttest answers to "Which biobrick controls only the efficiency - level of expression - of the genetic device?"

and the factors correlated to them. Most notably, the game was shown to correctly teach some of the Synthetic Biology content, be it vocabulary or mechanics, but this could not be generalized to online use. Indeed, students are affected by teacher and peer pressure to read through the text present in the game, while online participants are not and may skip through tutorials and explanations.

That is why we decided to conduct one additional experimental session of tests in another setting. We had noticed that classrooms made it impossible to separate students, inducing a group effect instead of a per-student effect. Online tests, on the other hand, had the drawback of not making it possible for us to detect when an issue arose, and to thoroughly monitor and analyze the subjects' behavior. A real-life playtest, involving subjects with whom we could interact, but who could not interact together was a more interesting setting for our experiment. This final playtest was conducted in April 2018 in the Cité des Sciences museum.

## 5.4 Final 2018 experiment at Cité des Sciences

An experiment was conducted in the Cité des Sciences museum in Paris in April 2018. The analysis of its results are presented in chapter 6. I chose this museum once more - the first public demonstration and the first playtest took place there in December 2013 and fall 2014 respectively - for several reasons:

- subject recruitment: the museum's staff could help us choose, recruit, and brief subjects, in a museum that attracts lots of very diversified guests - families, couples, youngsters, makers and so on;
- hardware: high-end gamer PCs are provided;

- troubleshooting: the museum's staff are accustomed to help scientists conduct experiments with the public, and in particular can help us in case of technical or human difficulty.

### 5.4.1 Objectives

We intended to have 100 subjects participate in the experiment, to match or exceed the size of the cohorts in other comparable studies. The review published by Connolly et al. (2012); E. A. Boyle et al. (2016) presents cohorts of sizes generally around 30-100. As in the previous campaigns, we wanted to assess the learning effect induced by the game, among other outcomes such as motivation.

Generally, we wanted to simulate online gamers' behavior. We wanted subjects to feel as free to play as they would at home.

Our secondary objectives consisted in getting general feedback about the game, to know of parameters that may impact the outcomes of the experiment. Preliminary experiments had already been conducted but we still wanted to know what the participants would report concerning the design of the game. We wanted to assess the understandability of the graphic interface, level design and game design, to evaluate the complexity of the content, and to list implementation bugs. This way, this final experiment could help us list up recommendations for improvements in the game.

### 5.4.2 Protocol

Participants were to be recruited without them knowing the full extent of the experiment, in order not to reduce the priming effect. Participants were to be told that the experiment consisted in a playtest, in which they had to:

- 1) fill in a pretest survey, in order for us to know more about their background. It would be more beneficial for us that they filled in additional, more detailed questions about biology, so that we knew about their prior knowledge in biology;
- 2) play the game at least 30 minutes, not necessarily finishing it;
- 3) fill in a posttest survey, to give some feedbacks about the game.

The "additional questions" of the pretest and the questions of the posttest (shown in annex C.4) were of course not really about the subjects' background, experience or feedbacks on the game, they helped us assess the increase of knowledge. The extended pretest and the posttest contain the same set of questions about the scientific content of the game. The participants did not know that they would be evaluated before and after playing, because in that case they would not have behaved casually, like online participants do. Either, they would have studied the game seriously, given up learning anything in the game right from the start,

or even given up participating in the experiment altogether - some people feel overwhelmed by the pressure of being assessed.

In case of an overabundance of guests, subjects were to be chosen in order to have a diverse cohort. It was expected that males aged 12-20 would be over-represented - as in the 2014 playtest -, so other demographic groups would be targeted first.

### 5.4.3 Implementation

We knew that the experiment presented some risks:

- having not enough subjects, which would make the experiment inconclusive.
- major technical difficulties, such as a blocking bug in the game, a faulty component on the PCs, or an unreliable Internet connection. This could strain the experiment pipeline, or even drive it to a halt, for instance in case of a complete loss of Internet connection.
- eventually unusable data, due to survey questions being ambiguous, due to a bug, or due to a bias.

To prevent this, we collaborated with the museum's staff to communicate with the public about the experiment. We tested the hardware and software beforehand the week before. In doing so, and in the first day of the playtest, we indeed encountered technical issues to which we found solutions. For instance, the network configuration of the museum prevented us from using the Internet version of the game, while the remote tracking events could be sent. We built a standalone version that we then installed on the PCs.

#### PC configurations

The configuration for each gaming station is given in the table A.2 on page 183.

This mainframe hardware configuration made the game run smoothly, without any lag or drop in audio or video frame rate.

The peripheral devices were also very good value and insured a good gaming experience. During the experiment, the headphones turned out to be our best asset in encouraging the subjects not to communicate while playing, to focus on the game, and to go on trying to solve the puzzles even when blocked.

#### Spatial organization of the experiment room

The Cité des Sciences could lend us one of two rooms in the Carrefour Numérique underground zone of the museum. Either a 30  $m^2$ , easily accessible and visible room without windows, or a much bigger, remote and hidden room with a big glass wall. We decided to opt for the more accessible and visible one. Moreover, the big glass wall could have been a source of distraction as it overlooked the park next to the museum.

We set up 5 gaming stations so that subjects could not watch each other's screen. We set up as few mobile walls as possible, as far as possible from the subjects, so that they would not to feel claustrophobic.

### **Timeline**

The experiment lasted 12 days over the span of 3 weeks from April 10th to April 28th, 2018. The experiment took place on Tuesdays, Wednesdays, Fridays, and Saturdays from 12:00 to 19:00 in order to match the time slots of maximum guest influx.

During the course of the experiment, as we gathered more subjects than what we intended for, we decided to use the last two days - April 27th and 28th - to do the same experiment but with a slightly different version of the game - see section 6.1.1.

After this presentation of the preliminary experiments and data gathering campaigns, and of the preparation and implementation of the final experiment, the next chapter presents and discusses its outcomes in regard to the research questions proposed in this thesis.

# Chapter 6

## Data analysis

In this chapter is described the process used to analyze data gathered in the April 2018 experiment presented in the previous chapter, in section 5.4. This analysis process is divided into two phases: an exploratory analysis, to search for overall patterns across the data and to check the validity of the gathered data, and a detailed analysis, to extract more precise information.

In most of this analysis, I used the python programming language associated on the Jupyter web application. Otherwise, I used tools such as simple graph-plotting in Google Spreadsheet.

### 6.1 Exploratory analysis

#### 6.1.1 Feedbacks from the participants

As a preliminary analysis, I focused on feedbacks given by the participants, including during the experiment, in order to detect problems early on. During the whole length of the experiment we gathered informal feedback from the experimental subjects by interviewing them when they were done answering the posttest. Their most frequent and meaningful remarks sorted by theme are presented in this section.

#### **The experiment**

The subjects mostly reported that they were positively surprised that such experiments were conducted in this museum. They did not expect such a level of interaction between the museum and scientists on the one hand, and between scientists and visitors on the other hand. They were also positively surprised by the content of the experiment: most of the visitors knew about game-based learning and citizen science games, but they did not know that there were such scientific *niche* games. This tends to support the acceptability of such games and their prospects of growth.

### The experimental setup

The most common first reaction of the subjects was to notice the quality of the computers lent to us by the Cité des Sciences. Youngsters especially noticed that they were to play on high-end gamer PCs.

The most frequent negative remarks dealt with the lack of space in the experiment room. Some even reported that we distracted them when we walked around the gaming stations to make sure that the experiment was running as planned.

### The game

Similarly to the previous experiments, there were different typical reactions to the game:

- subjects that either had an education in biology or were typically older than 40 wanted more information to be available within the game, worrying that gamers with no education in biology could not play or could not understand what was going on in the game. They usually had a low knowledge and practice of games.
- subjects that had no education in biology but had some knowledge of video games appreciated the graphics above all, followed by the music and the gameplay. They had a great time playing nonetheless but did not necessarily understand what was going on nor remember most of the technical names of the BioBricks and of other game elements. They usually asked for less information and text to be displayed, or at least for a simplification of the names and mechanics.

There was of course a wide range of reactions to the game, most of them positive, but in some rare cases the very principle of mixing education with video games was criticized.

As said in section 5.4.3, we decided to implement a series of improvements before the end of the experiment. Indeed, we noticed that the subjects would frequently report their struggling to grasp a number of notions or to complete certain chapters. Measuring or estimating the impact of the adjustments we introduced would be of interest to this study as it would help to show how to increase the impact of the game through the chosen game-based learning strategy implemented in Hero.Coli, stealth learning. Indeed, stealth learning uses intrinsic motivation by embedding the learning content in the gameplay without breaking the flow; if a change in gameplay triggered a change in knowledge acquisition, or at least in behavior and appreciation of the experiment, it would confirm that the more gameplay is tuned and tweaked to facilitate progression in the game, the more learning and motivation are fostered. Moreover, two of these adjustments were specific level design modifications which would not only make the game slightly easier, but most importantly more *consistent*. Consistency contributes to the immersion of the participants, and consequently increases the outcomes of playing

the game on motivation and learning. The other adjustments were not as critical but solved confusing situations. We hypothesized that all of these adjustments would bring more positive impacts through an increased immersion than negative impacts through making the game less challenging, challenge being another factor that increases the outcomes of playing video games through increased engagement. We deemed these improvements as a way to get a more balanced difficulty level overall.

The following is a list of previously implemented challenges, game mechanics, and feedbacks which were reported as confusing by the participants:

- Counterintuitive, blocking puzzles: in two locations of the game shown in figure 6.1, the player had to solve the same kind of counterintuitive puzzles.

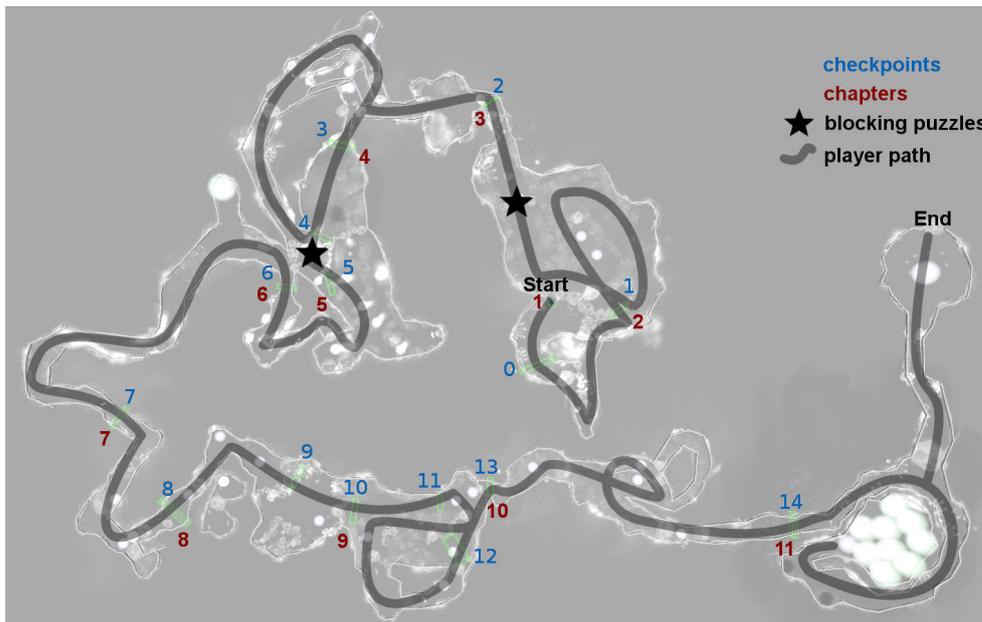


Figure 6.1: Map of the game with highlighted the two blocking puzzles

They had to go against an earlier tutorial recommendation by swimming across an antibiotic wall. The speed of the avatar made it possible to cross the toxic cloud without dying, but that mechanic was not clearly established: this was not explicitly given as a mission, even though it was *hinted* once when two Non-Playing Characters (NPCs) demonstrated it in a cutscene (an in-game demonstration video). Therefore, most of the players did not know what they had to do. Moreover, this antibiotic-crossing mechanic had no educational value, and would lead the players into swimming in tedious time-consuming loops, early on in the game. The solution we chose was to replace those counterintuitive antibiotic crossings by simple door-like elements. These elements, when pushed, explicitly describe the conditions on which they can be opened if they are not already fulfilled, therefore fixing the consistency problem and giving directions to the players.

- Deceiving, unclear puzzle: similarly, another puzzle could be blocking later on in the game. In chapter 5, there is no ambient light, meaning that the bacterium swims almost in complete darkness. The player has to use an arabinose-induced GFP device to spot the traps. But arabinose clouds were not visible, even when the GFP was excited. Therefore, some players had to work by accident, or went through a painful workaround. Even worse than that, to complete this chapter, the player had to use a non-arabinose-induced RFP device. This, compounded with the green-blue red-green confusing pattern listed above lead to some players being stuck, or blindly solving the puzzle by randomly assembling BioBricks. We solved this by making the puzzle unambiguous and consistent: the arabinose clouds are more visible and an arabinose-induced RFP device is required to complete the chapter.
- Fluorescence mechanics: the green fluorescent protein (GFP) and red fluorescent protein (RFP) are available in the game. The GFP, cited in sections 2.1.1 and 2.2.3, emits green light when excited by blue light and the RFP was then erroneously depicted to emit red light when excited by green light, when it is actually excited by orange light. This green-blue red-green pattern would greatly confuse the players. Fixing this error increases the scientific accuracy of the game and is likely to make the fluorescence mechanics less confusing.
- Insufficient feedback on DNA strand pick-up: in the previous version of the game, picked up DNA strands would only display an explanatory pop-up the first time they were picked up. This means that the next times they were picked up they would only trigger an audio feedback. We observed that in that case very few participants immediately stopped swimming to look at their new BioBrick. This would hinder the progression of the participants: as they only saw the explanatory pop-up once per BioBrick, they did not really know which and how many BioBricks they had and what these BioBrick could do. Displaying the pop-ups at every pick-up would contribute to cope with this issues.
- Purely decorative scenery elements: some non-interactive ambiguous lights and creatures made the expected behavior difficult to guess. For instance, some lights looked like GFP and RFP excitation lights; some creatures looked like elements to pick up. To solve this problem I edited, moved, or deleted these scenery elements, and emphasized even more some interactive elements.
- Checkpoints: players complained about the fact that their avatar bacterium collided and bounced off the save point bacterium. This save point bacterium appears like a ghost and is the point from which they will respawn and play again after having died. In most games the player avatar does not bounce off those ghosts, if there are such ghosts. In the game we decided to truly materialize them to show that they are not ghosts, but real bacteria produced by division. But due to the complaints of the players, and the fact that in

some cases it could even get the players stuck, we decided to make them non-obstructing.

- Frustrating map design: some parts of the level are like hubs or loops (see figure 6.1). It is disturbing for some participants, especially the non-gamers, who think they are lost. This also primes the participants into getting lost later on in the game, when they try looping in areas were they are not supposed to loop. This qualitative observation has yet to be evaluated quantitatively in terms of frequency of occurrence and of detrimental effects on game completion and learning. This could be done in further studies by comparing the Hero.Coli version used in this experiment with another very similar version that does not feature loops.
- Frustrating level design: some game elements behaving like anti-personnel mines had a poor placement that frustrated players. We moved them to keep the gameplay while still being playful.
- Implementation bugs: the behavior of some doors was affected by a bug that could, in some rare cases, lead to a dead end or to a shortcut in the story.

These fixes had the additional benefits of being easy, quick, and not hazardous to implement.

### **The pretest and posttest surveys**

Several participants were confused by the pretest and posttest surveys. Some were confused by the general idea, others by the fact that in our implementation the pretest and posttest are technically the same Google Form document. The pretest and posttest are two different paths inside the same survey, branching from the question "Have you ever played Hero.Coli?".

This pretest-posttest confusion, besides being voiced by some experimental subjects, had also been reported in previous experiments. As described in section 5.1, we had chosen to merge the two surveys in order to minimize development costs, assuming that the pretest-posttest switch would not be confusing. The feedback we got showed that it was, for some subjects.

Having presented the qualitative analysis of the feedbacks of the participants to the experiment, the next section presents the quantitative analysis of survey and tracking data. To get a measure of the outcomes of the game, I had to compare exploitable pretest and posttest results. Using the surveys and the remote tracking, I first filtered out unusable or irrelevant data. The user ID described in section 4.3.2 made it possible to easily associate tracking data with surveys.

### 6.1.2 Filtering the data

The game was still playable online during the experiment, therefore survey results and remote tracking data contained data from online participants and from participants to the Cité des Sciences experiment, and had to be filtered.

89 subjects out of the 193 participants could be used in the first phase of the study. This shows that due to various reasons only a portion of the cohort can be used. This will reduce the significance of the results and may even prevent the analysis to be conclusive on a set of research questions. In order to compute that number, the data were first filtered according to their date and according to the platform the game was launched on. Both these pieces of information are accessible in the tracking data. Only exploitable data were kept. I used as criterion the amount of game events sent using RedMetrics and the existence of chronologically correct pretest-posttest pairs bound to these RedMetrics events. Finally, data of participants who only filled in the restricted set of compulsory questions, and data affected by technical errors were discarded. Figure 6.2 shows how much data was discarded for each reason, starting from the pool of identified unique participants matching the platform and time slots of the experiment.

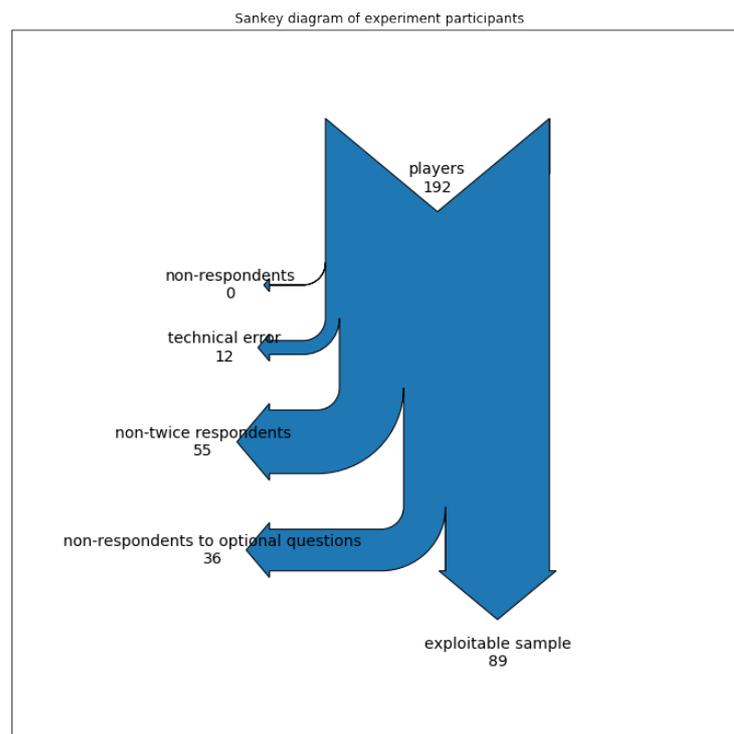


Figure 6.2: Pipeline of exploitation of the data from experimental subjects

Those reasons for only 89 sets of data out of 192 being exploitable were identified as most likely being:

- the protocol not being respected:

- insufficient preparation and continuous monitoring by the experimenters, while too much control would interfere with the experiment by introducing various biases (priming, stress, ...).
- the insufficient possibility for visitors to give feedbacks about their misunderstanding or discomfort with the protocol.
- the protocol being impractical for some visitors: too long, too tiring for a holiday.
- technical issues that for instance led the subjects to not fill in the posttest, or to play another version of the game.
- technical issues:
  - failed to create a new identifier for some visitors.

### 6.1.3 Correlations in survey answers

To answer my research questions pertaining to demographics, relationship to games and biology, scores, I studied the correlation between them, by plotting tables of correlations. These tables of correlations are not to be confused with correlation matrices: a correlation matrix is a symmetrical matrix in which element  $(i,j)$  is the correlation between  $X_i$  and  $X_j$ , while in the tables presented in this section, rows and columns do not contain the same features, *i.e.* element  $(i,j)$  is the correlation between  $X_i$  and  $Y_j$ . Tables of correlations are used here instead of correlation matrices in order to enhance the legibility of the results.

#### Participant characteristics and score

Figure 6.3 is the table of correlations between scores and participants' self-assessed data.

An enlarged and annotated version of this table is available in annex B.4. This figure makes it possible to discuss my research hypotheses.

#### Application to research hypotheses

- **H3a: "age correlates negatively with learning outcomes; gender does not"**: this table clearly demonstrates the absence of correlation between age and scores (posttest score and score increase) and gender and scores.
- **H3c: "learning outcomes correlate positively with interest and education in biology"**: curiosity about biology and synthetic biology, education in biology are positively, moderately correlated to posttest score and score increase. Interest in biology is only weakly positively correlated to posttest score and score increase.
- **H3e: "learning outcomes correlate positively with interest and practice in games"**: there is no evidence of such correlation.

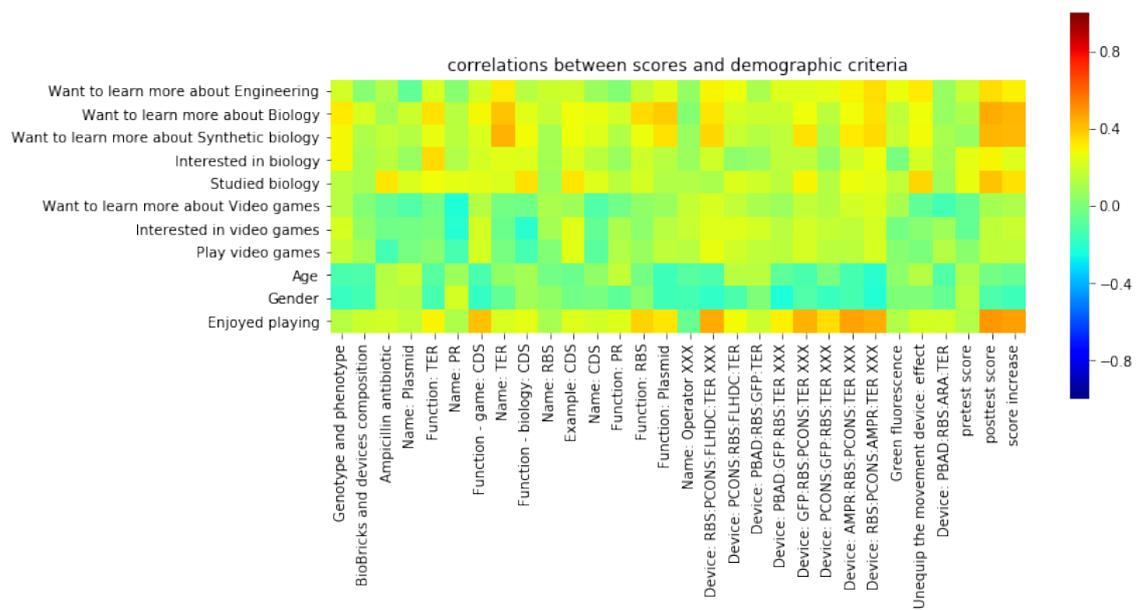


Figure 6.3: Table of correlations of demographic features and interests against scores

### Participant demographic data, curiosity, interests, practice, enjoyment

Figure 6.4 is the table of correlations between participants' self-assessed demographic data and curiosity, interests, practice of biology and video games (enlarged version in annex B.5). This figure shows the absence of correlation between age and scores and gender and scores. This answers the research hypothesis **H3b: "age correlates negatively with motivation; gender does not"**.

Figure 6.5 represents the correlations between enjoyment and the other characteristics of participants (enlarged version in annex B.6).

This figure shows several correlations: enjoyment (the proxy for engagement) is moderately positively correlated with curiosity in engineering, biology, and synthetic biology, with interest in biology and video games, and with the practice in video games. No correlation was found with education in biology, and only a weak correlation was found with curiosity in video games, and age and gender. This partially validates research hypothesis **H3d: "motivation correlates positively with interest and education in biology"** and validates research hypothesis **H3f: "motivation correlates positively with interest and practice in games"**.

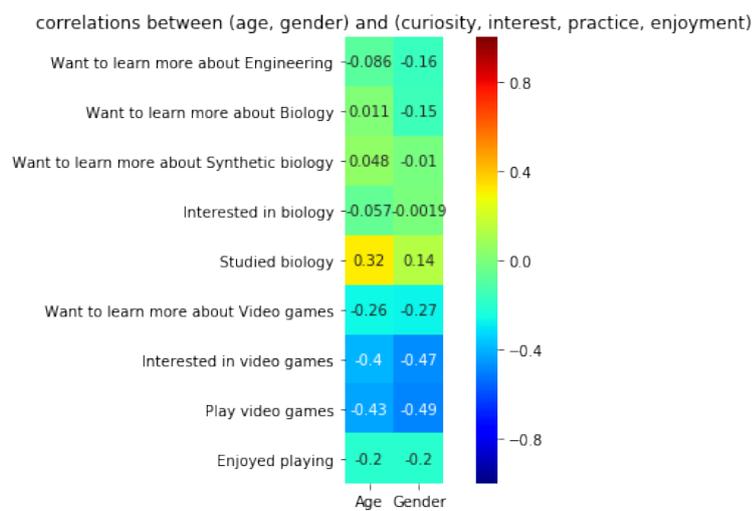


Figure 6.4: Table of correlations of participant demographic features against their curiosity, interests, and practice

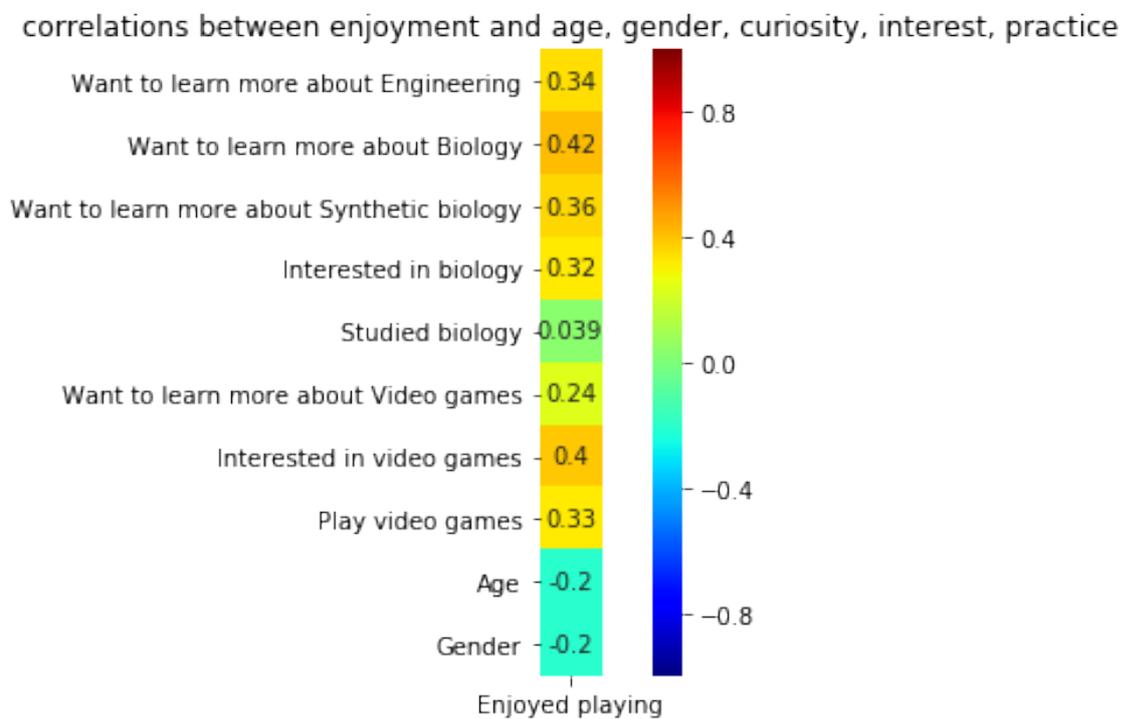


Figure 6.5: Table of correlations of enjoyment against participants' characteristics

### 6.1.4 Correlations between tracking data and surveys

I used the same method as in section 6.1.3 to answer my research questions pertaining to game metrics and self-assessed participant characteristics. I compared participant characteristics and duration of play, then scores and duration of play.

#### Participant characteristics and play times

Figure 6.6 is the table of correlations between play times and participants' self-assessed data (enlarged and annotated version in annex B.7)

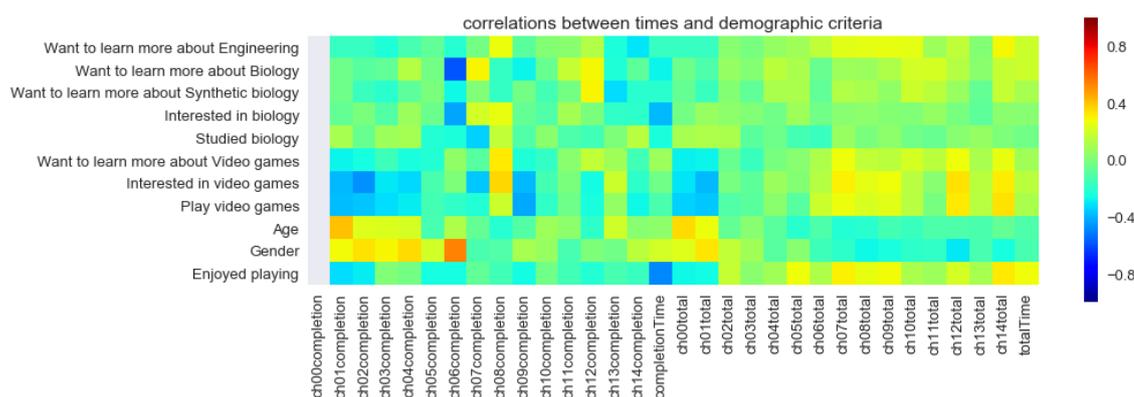


Figure 6.6: Table of correlations of play times against participants' self-assessed data

This figure only partially confirms hypothesis **H4a: "age correlates positively with playing duration to complete the game; gender does not"**: age does correlate positively only for completion times and total spent times in the first sections of the game; gender correlates positively in a similar fashion. This may be due to a sampling bias linked to the small count of females in the sample: 24 as shown later in section 6.1.5. It also partially confirms hypothesis **H4b: "playing duration to complete the game correlates negatively with interest and education in biology"**: the total completion time correlates negatively with interest in biology, but not with curiosity or education in biology. Finally, hypothesis **H4c: "playing duration to complete the game correlates negatively with interest and practice in games"** is also partially confirmed. Interest and practice in games correlate negatively only with completion times of the first four sections, and with total times in the first two sections. They correlate positively with total spent times in the last sections, possibly due to a selection bias: gamers may be more prone to being able to simply access these end-of-game sections. Curiosity in video games does not correlate with playing times.

The biggest correlations are between completion time of checkpoint 6, and curiosity in biology and gender, with a moderate negative (respectively positive) correlation factor.

## Play times and scores

Using remote tracking data makes it possible to measure the time spent by a player on a section of the game, be it the total time spent, or the time to complete the section. For instance a player may go back to a section to search for clues or missed items, or the player may start again from an earlier section. Figure 6.7 (enlarged and annotated version available in annex B.8) is a table of correlations of the completion times and total times per section delimited by checkpoints 0 to 14 (see figure 4.12 for their position on the game map) against each game content question, the pretest score, the posttest score, and the delta score, *i.e.* the score increase (posttest score - pretest score).

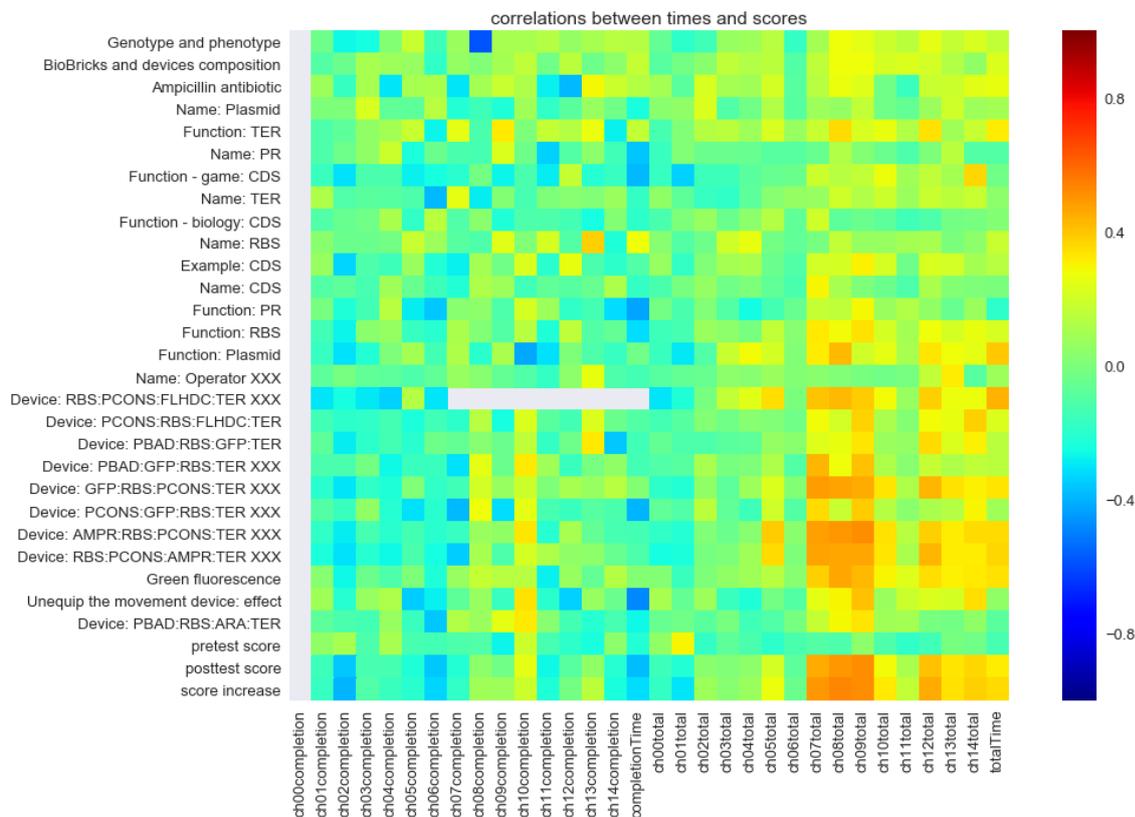


Figure 6.7: Table of correlations of the play times against score per question and total scores

All of the questions are replaced by codes to increase legibility. The list of questions and of their associated codes is given in annex table A.3. The codes are designed to be shorter and to contain additional information about the context of the question and about the answer: code prefixes ("Device", "Name", "Function") sum up the question, codes containing "XXX" mean that the answer is "None" (because the DNA sequence is not functional for instance), codes containing device codes describe the image given in the questionnaire. The complete questionnaire

as presented to the participants is given in annex C.4.

Negative correlations are in blue and positive correlations in orange. White pixels show either absence of data or undefined correlations, due to no variation on the answers given by participants or on the time spent by participants on this section. The absence of data is explained by figure 6.8: it shows on how many participants each correlation is based. They represent the number of participants who have reached this checkpoint for the top half, because the time needed for completion for a checkpoint that was not reached would have been meaningless. The total time spent around checkpoint  $n$  by participants who did not reach it is 0.



Figure 6.8: Number of participants on which the correlations of figure 6.7 are based

Going back to figure 6.7: on the y-axis are completion times and total times. The completion time of checkpoint  $N$  is the elapsed time between the time of completion of the checkpoint  $N-1$  and the first time the checkpoint  $N$  was reached. The total time of checkpoint  $N$  is the time spent playing nearby checkpoint  $n$ , *i.e.* between checkpoint  $n$  and  $n + 1$  when moving forward and between checkpoint  $n$  and checkpoint  $n - 1$  after a U-turn.

The left half of the graph, featuring completion times, presents a general weakly negative correlation trend, with noise. This phenomenon is more visible in the left-side quarter of the graph, featuring completion times for checkpoints  $n \leq 7$ . This could be explained by the fact that players who presented more difficulties in the beginning of the game tended not to finish the game and therefore had less access to its learning material. The difficulties they encountered may be linked to usability and acceptability issues (notably those presented in section 6.1.1) rather than the scientific content of the game itself. The "completion time" row, sum of all checkpoint completion times, is mostly negatively correlated to scores. This means that, overall, spending too much time completing each checkpoint is negatively correlated with correct answers in the posttest, and thus to the posttest score. Checkpoint 10 is the only one presenting a consistent slightly positive correlation to device-related questions. In this section between checkpoint 9 and 10, the player can get an additional optional item if they spend time solving a puzzle, but they cannot access it anymore once they reached the checkpoint 10. Therefore, more curious players will spend more time trying to get this optional item, and consequently will spend more time reaching checkpoint 10. This positive correlation would be a hint that curious players tend to score higher in this set of

questions.

On the right side, a consistent pattern of moderate correlations (orange regions) shows that the total time spent around checkpoints 7 to 10 and 12 to the end of the game positively correlate with better posttest answers in a group of approximately 6 questions, and thus with the posttest score. More generally, the total times for sections after checkpoint 6 generally correlate positively with a 0.3 coefficient (weak-moderate correlation) with the score on most of the questions. This right-side quarter of the graph with higher correlations could then be explained by the fact that players who spent time in those later sections of the game at least *had to reach* those later sections and therefore have a certain mastery of the game mechanics. A noteworthy and easily explainable exception is the spent time around checkpoint 11: there is no correlation at all with any score. The area in the game around checkpoint 11 is indeed small and straightforward, inducing very little variation over the time spent to traverse it. The remark I made in the previous paragraph about checkpoint 10 still stands, and is even more visible there: the more players spent time between checkpoint 9 and 10 in the forward progression sense (it is actually impossible to go backwards from checkpoint 10), the more they answered correctly and had a greater score, which confirms the link between curiosity and score.

When looking at this graph along the questions axis, we can see that the pretest score does not correlate with any play time. This is not surprising as the pretest scores have very low mean and standard deviation (see table 6.3). We can also see that the questions that correlate the least in absolute value are linked to vocabulary: question codes prefixed with "Name", and the biological interpretation of the Coding Sequence ("Function - biology: CDS"). This would mean that the speed and total time spent in the game do not correlate with the memorizing of vocabulary. Finally, the questions that correlate best with the total time spent in the various sections of the game are four questions about BioBrick assembly order (starting with "Device"), the question about green fluorescence, and the plasmid question. They are all linked to game mechanics tied to frequently-used visual elements, especially in the second half of the game.

### 6.1.5 Surveys: Analysis of the cohort

The cohort obtained from the filtering process was split up according to declared age, gender, interests, education in biology, and practice in video games. The results are given in the following pie charts, figure 6.9, 6.11 for gender and age respectively.

Additionally, I defined what "positive answers" are for the following types of questions, for a quicker reading of results:

The answers to game content questions were counted as positive when they matched the unique correct answer to each question.

For questions relative to the demographic group I arbitrarily counted as positive Female, age group 18-25, English language. In principle none of those questions admits a binary answer, it is merely indicative.

For questions relative to the interests, practice, and experience:

The following answers to the questions pertaining to interests and game practice were counted as positive: "Moderately", "A lot", "Extremely". This choice is biased in favor of positive answers, as there are three options out of five possible subjective answers.

The following answers to the question "How long have you studied biology?" were counted as positive: "Until bachelor's degree", "At least until master's degree". This choice is biased in favor of negative answers, as there are two options out of five possible answers that are partially subjective - the participants who are still studying may interpret "studying until degree X" as having obtained it, or not.

The positive answer to the question "Have you ever heard about synthetic biology or BioBricks, outside of Hero.Coli?" is "Yes, and I know what it means".

The following definitions will be used to refer to groups of participants:

"Biologists" are respondents who answered *positively* as defined above in at least one of the four following questions: "How long have you studied biology?", "Are you interested in biology?", "Before playing Hero.Coli, had you ever heard about synthetic biology?", "Before playing Hero.Coli, had you ever heard about BioBricks?".

"Gamers" are respondents who answered *positively* as defined above in at least one of the two following questions: "Are you interested in video games?", "Do you play video games?".

#### Gender

For comparison, the proportion of genders of the 407 online participants on the 2018-07-05 - 2018-09-19 period is shown in figure 6.10. Among these 407 online participants, 88 reported their gender as female, 249 as male, 15 as other, and 55 did not specify any. I speculate that the main difference - there are far more other and unspecified genders online - may be explained by the bias introduced by the presence of experimenters near the participants, a problem I tried to cope with while designing the questions - see section 5.1. Online participants may feel more

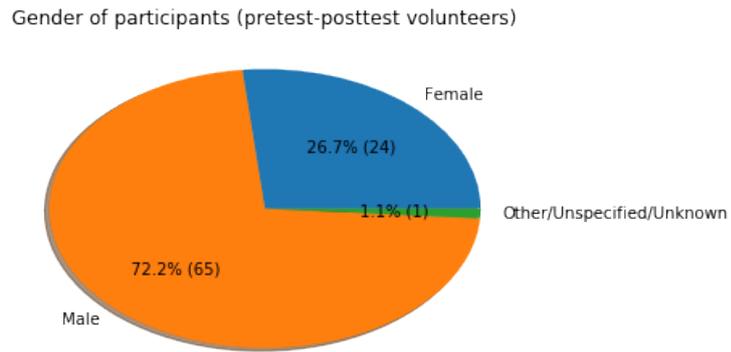


Figure 6.9: Gender of the participants kept in the study

free to choose, while the anonymity also enables other participants to deliberately enter incorrect information. This can be verified in the ages entered by online participants: 2 people reported that they are 120, 3 reported that they are 0-1, and so on.

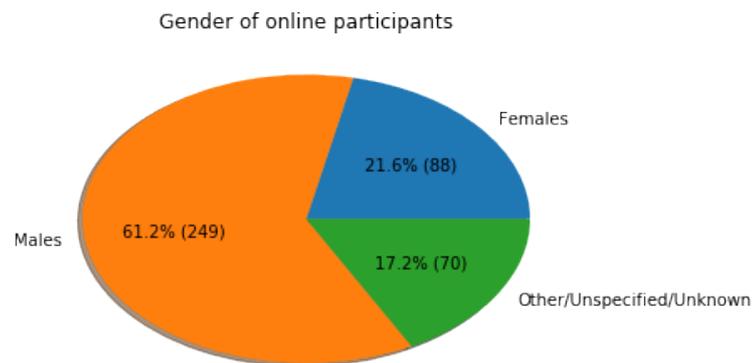


Figure 6.10: Gender of online participants (407 people, 2018-07-05 - 2018-09-19 period)

## Age

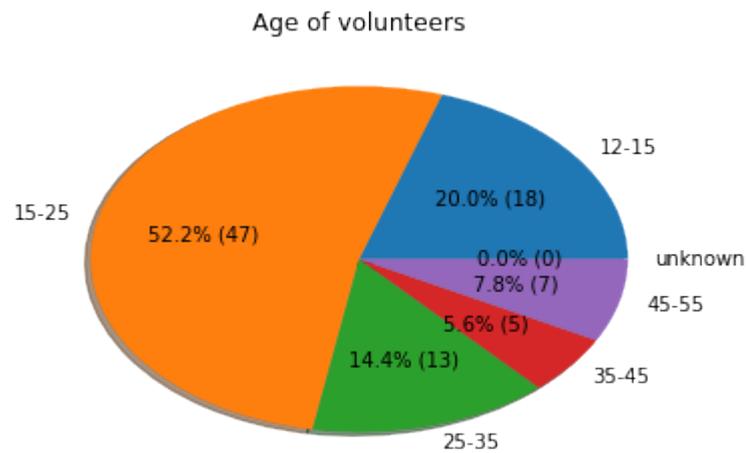


Figure 6.11: Age of the participants kept in the study

The cohort is diverse. The 10-25 and male classes are overrepresented compared to the French population. In terms of in-class use, though, the age overrepresentation is not an issue. It does not completely match the age repartition of online participants: online participants are underrepresented over 45 and under 15 years of age - see figure 6.12. But there are approximately 11% of intentionally incorrect reported ages, because ages above 90 or under 12 can be assumed as unrealistic.

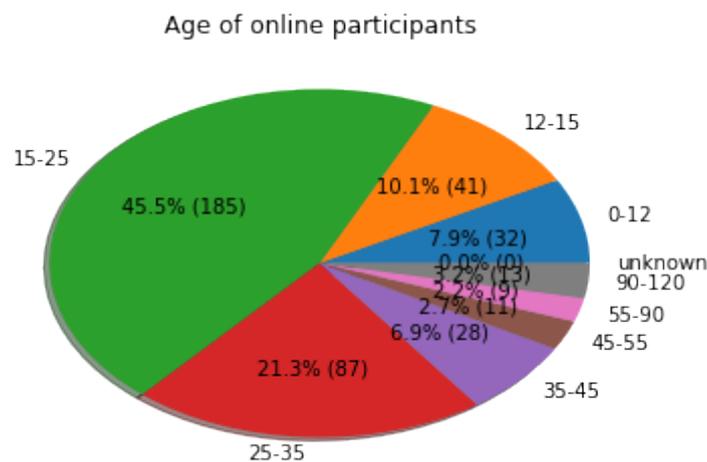


Figure 6.12: Age of online participants (407 people, 2018-07-05 - 2018-09-19 period)

## 6.2 Detailed analysis

### 6.2.1 Threshold effect: learning and checkpoints

Two checkpoints were identified as thresholds in the learning process: mostly checkpoint 5 and more marginally checkpoint 2 represented in figure 6.1. It means that participants who reached those checkpoints and beyond gave more correct answers than those who did not reach those checkpoints. However, not all questions have thresholds, and not all the questions have the same thresholds. To show that, the furthest checkpoint reached by each participant was computed from the tracking data, and correct answers were counted in the survey data. The resulting bar plots are shown in figures 6.13, 6.14, and 6.15. Figure 6.13 shows a case without any obvious threshold because most of the answers are incorrect at any given checkpoint. Similarly, figure 6.14 is a case without any obvious threshold because most of the answers are correct at any given checkpoint. Finally, figure 6.15 shows a situation with a threshold: checkpoint 4 or 5 depending on the ratio of correct answers set to define a threshold.

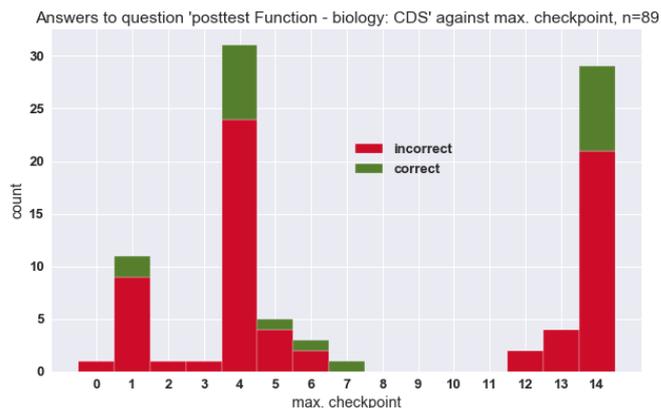


Figure 6.13: Posttest answers to question 14, subquestion 6

A checkpoint will be defined as a learning threshold for a question if a ratio  $R$  of the answers given by the participants who reached at least this checkpoint is correct. For instance, in figure 6.15 with  $R=75\%$ , checkpoint 5 will be a learning threshold for question 15, subquestion 11 if 75% of the participants who went further than checkpoint 4 answered correctly to this question. To get rid of this dependency on the ratio  $R$ , different values of the learning threshold were plotted against  $R$  to show those which are stable across ranges of values of  $R$ . Likewise, some learning thresholds are stable and some are not. In figure 6.16, checkpoint 5 is a clear learning threshold due to the plateau, while in figure 6.17, there is no clear plateau. To automatize in part the analysis, for each question the most frequent value of learning threshold was computed on the value range of  $R$  on  $[0.5, 1]$  using 11 value points increasing by steps of 0.1. Learning thresholds showing

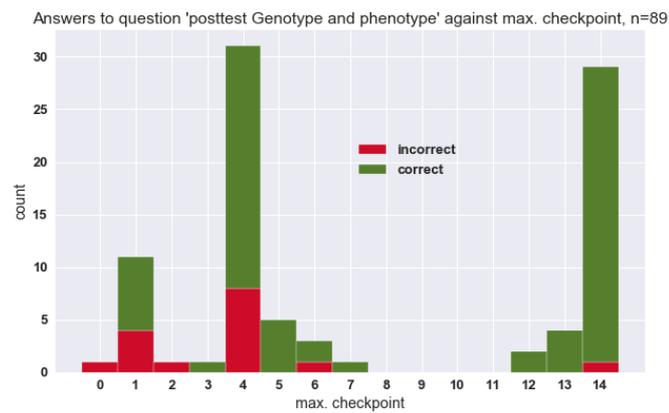


Figure 6.14: Posttest answers to question 11

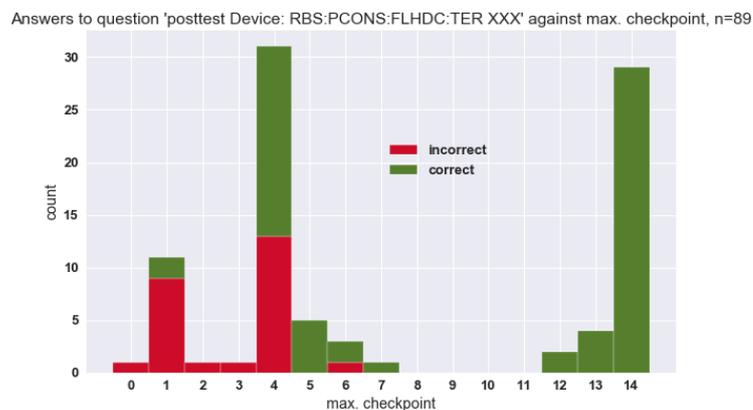


Figure 6.15: Posttest answers to question 15, subquestion 1

that too few or too many answered correctly certain questions were discarded, and only the thresholds that were repeated at least twice and effectively had a plateau (stable value in the neighborhood, steep increase before and after) were kept. The results are gathered in table 6.1.

This threshold effect does not appear for the total posttest score. Several participants who did not go further than checkpoint 4 got a high score while several participants who completed the game got a low score. To detect a threshold effect I used segmented regression (figure 6.18). It shows an inflexion on checkpoint 7 and very similar steepnesses for the regression before and after checkpoint 7. Likewise, a linear regression yields acceptable results (figure 6.19). Its parameters are presented in table 6.2.

This section answered **RQ6: Quiz-based assessment and automated tracking "H6a: In a linear game where each puzzle is compulsory to**

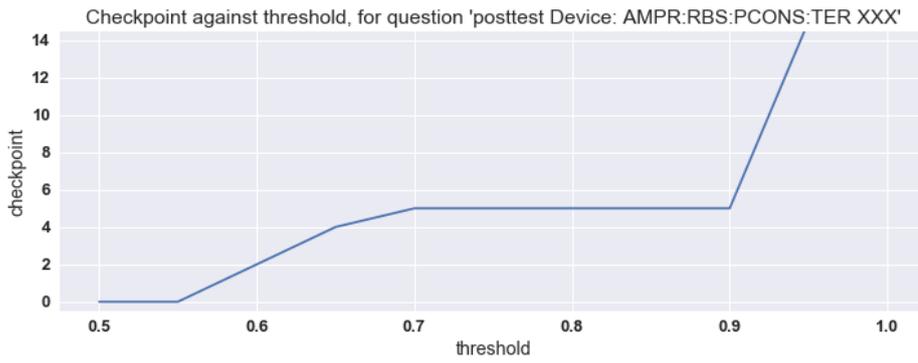


Figure 6.16: Learning threshold vs ratio criterion for question 15, subquestion 7

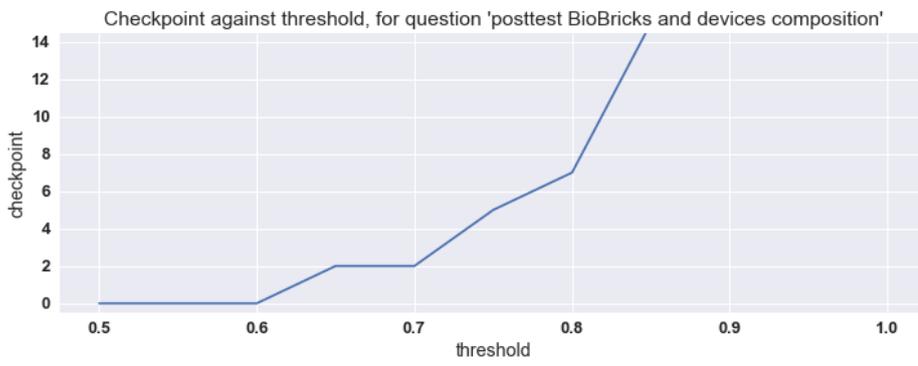


Figure 6.17: Learning threshold vs ratio criterion for question 12



Figure 6.18: Total posttest score vs furthest checkpoint reached

solve one after the other to finish the game, reaching some thresholds

Question	Learning threshold
Q11Genotype and phenotype	-
Q12BioBricks and devices composition	-
Q13Ampicillin antibiotic	5
Q14-01 Name: Plasmid	-
Q14-02 Function: TER	5
Q14-03 Name: PR	-
Q14-04 Function - game: CDS	-
Q14-05 Name: TER	5
Q14-06 Function - biology: CDS	-
Q14-07 Name: RBS	-
Q14-08 Example: CDS	-
Q14-09 Name: CDS	-
Q14-10 Function: PR	-
Q14-11 Function: RBS	5
Q14-12 Function: Plasmid	5
Q14-13 Name: Operator XXX	-
Q15-01 Device: RBS:PCONS:FLHDC:TER XXX	5
Q15-02 Device: PCONS:RBS:FLHDC:TER	5
Q15-03 Device: PBAD:RBS:GFP:TER	-
Q15-04 Device: PBAD:GFP:RBS:TER XXX	5
Q15-05 Device: GFP:RBS:PCONS:TER XXX	5
Q15-06 Device: PCONS:GFP:RBS:TER XXX	-
Q15-07 Device: AMPR:RBS:PCONS:TER XXX	5
Q15-08 Device: RBS:PCONS:AMPR:TER XXX	5
Q16Green fluorescence	5
Q17Unequip the movement device: effect	5
Q18Device: PBAD:RBS:ARA:TER	-

Table 6.1: Learning thresholds checkpoints associated to survey questions

Intercept p-value	$3.8 * 10^{-7}$
Slope p-value	$2.1 * 10^{-12}$
$R^2$	0.43

Table 6.2: Linear regression parameters for posttest score against furthest checkpoint reached

can be equivalent to validating the assimilation of a notion." and **RQ7: Threshold effect H7a:** "After reaching a certain point in the game, most of the learning is achieved: going further will not increase the learning outcomes.". Indeed, certain questions indicate checkpoint 5 as a learning threshold (H6a), but this does not transfer to the overall posttest score, negating H7a.

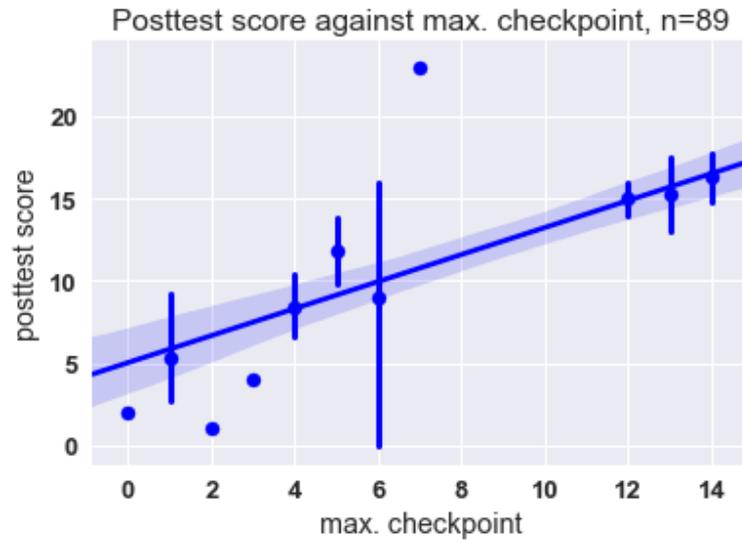


Figure 6.19: Total posttest score vs furthest checkpoint reached

### 6.2.2 Comparison of the pretest and posttest pairs

#### Variation of correct answers

Figure 6.20 and figure 6.21 show percentages of positive answers in pretest, posttest, and change. Figure 6.20 presents the results using the order of questions of the survey, and figure 6.21 presents the results by increasing variation between pretest and posttest. Both these graphs are also available in enlarged versions in annex B.3.

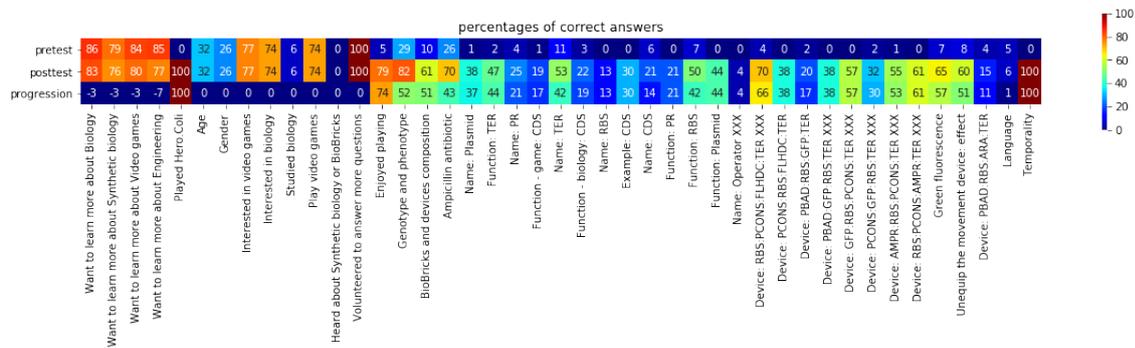


Figure 6.20: Percentages of positive answers in pretest, posttest, and percentage increase

These graphs show an increase in correct answers for the majority of questions - only two questions out of the 27 game content questions show an increase smaller than typical statistic noise in this study (around 10 percentage points).

This increase is not uniform over all the answers. Questions pertaining to

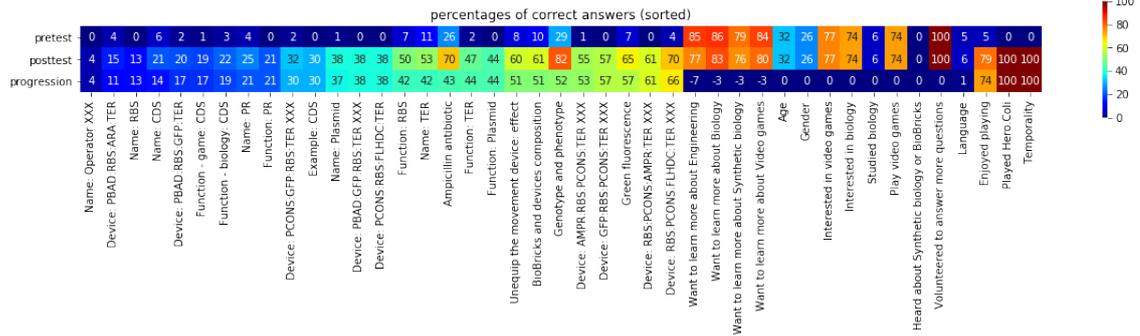


Figure 6.21: Percentages of positive answers in pretest, posttest, and percentage increase, sorted by increase

vocabulary - the questions which have a code prefixed with "Name" show poor results in posttest and progression scores. This confirms one of the conclusions of section 6.1.4: the vocabulary of the game is not mastered by the players. On earlier versions of the game, which had less tutorials and pop-ups, this vocabulary was however learned by students who played the game in-class (see section 5.3): this may show that students in a school context are more engaged than participants in a public museum. Section 6.1.4 also showed that the correct answering on the question about ampicillin did not correlate with time spent in the game. Here we can see, however, that the score increase and posttest score on this question are sizable - 43 percentage points and 70% respectively. It shows that, generally, however long the participants played, they remembered the name "ampicillin".

Questions pertaining to incorrect BioBrick assembly, green fluorescence, nature of BioBricks and devices, effect of equipping genetic devices have the best progression and posttest scores - with the exception of question "Device PCONS:GFP:RBS:TER XXX" which was answered correctly by "only" 32% of participants, with an increase of 30 percentage points. Questions pertaining to BioBrick vocabulary, BioBrick functions, complex induced device function have the lowest progression and posttest scores. BioBrick function is very unevenly mastered: the promoter function is known by only 8% of participants, CDS, 18% (game function) and 22% (biological function), terminator, 47%, and RBS, 50%.

These conclusions must be put into perspective using the phenomenon discussed in section 6.1.4: the correct answer rate on several questions is correlated with the time spent in some sections of the game. Some questions may have a low overall progression and posttest score simply because most players were stuck in earlier parts in the game. For instance the questions with a code prefixed with "Device: PBAD", related to the advanced concept of inducible promoters described in section 2.2.2, have very low overall scores and score progression, but their posttest and progression scores correlate positively with time spent in the later sections of the game. Therefore, the solution to increase the score on some questions may be to facilitate the access to the later parts of the game, in other words, to make the

learning more progressive or adaptive. It might also just be a sample bias: players who are "selected" by the game are the ones who had the prior knowledge required to understand inducible promoters, for instance, while the game itself might not be efficient at teaching this notion.

**Application to research hypotheses** This analysis makes it possible to discuss **hypothesis H2b: "the use of a game can make citizens understand basic notions about synthetic biology"**.

- a) "the simplified link between genotype and phenotype": 82% of participants answered correctly in the posttest to the question pertaining to this link, of code "Genotype and phenotype", with an increase of 53 percentage points from pretest.
- b) "the nature of BioBricks and devices: DNA sequences": 61% of participants answered correctly to the related question in the posttest, of code "BioBricks and devices composition", with an increase of 51 percentage points.
- c) "the BioBrick simplified grammar (Promoter - RBS - Coding Sequence - Terminator)": as discussed above, questions pertaining to incorrect BioBrick assembly are among the best concerning progression and posttest scores.
- d) "the names and functions of the bricks (condition - quantity - function - end)": as discussed above, questions pertaining to BioBrick names are among the worst in terms of progression and posttest scores. The questions pertaining to brick function have slightly better results.
- e) "the advanced notion of inducible promoters": as discussed above, the questions related to induction have low overall rate of correct answers. This may be linked to the fact that only a minority of players played all the sections of the game dealing with it.

**Anomalies** As a side note, I found some answers to be inconsistent in figure 6.20: some participants changed language between the pretest and the posttest, and some answered "Yes" in the pretest when asked whether they enjoyed playing Hero.Coli.

### Direct comparison

The paired t-test between the pretest and posttest groups on the questions pertaining to the content of the game is consistent with these groups being statistically different: t-value=  $-13.94$ , p-value=  $3.03 * 10^{-30}$ .

score metric	pretest	posttest	progress
mean score	1.36	11.2	9.88
median score	1.00	12.0	11.0
standard deviation	1.69	6.46	4.77

Table 6.3: Mean, median, and standard deviation on pretest, posttest, and delta scores

When comparing subgroups, the following p-values are obtained, which show the same results for subgroups - see table6.4.

category	group size	p-value
all respondents	89	$3.0 * 10^{-30}$
females	24	$1.3 * 10^{-7}$
males	64	$2.3 * 10^{-24}$
biologists	0	-
gamers	66	$7.4 * 10^{-24}$

Table 6.4: T-tests between pretest and posttest scores among some player groups

### Finer score analyses

Earlier experiments using earlier versions of Hero.Coli demonstrated that learning happened but that misconceptions could be introduced (see section 5.3).

In order to assess the level of understanding and show the existence of misconceptions in this version of the game, I first labeled the possible answers for each question according to their accuracy: "severe misconception", "mild misconception", "does not know", or "correct". I then used two methods: for single questions, I directly compared the number of answers with the different labels in the pretest and posttest surveys. I used Sankey diagrams to represent the flows between the different answers. For groups of questions, I used different *gradings* *i.e.* weights on the different types of answers. These gradings are shown on table 6.5. The grading used until now - labeled *base* - gives 1 point per correct answer and 0 point otherwise. In all the other gradings, the answer "I don't know" gives more points than answers related to a severe misconception. The *lenient* grading does not penalize mild misconceptions as much as the *base* grading, even giving more points to mild misconceptions than to the "I don't know" answer. The *category-lenient* grading does not distinguish between a correct answer and a mild misconception that is at least pertinent, of the same category of objects. The *strict* grading penalizes misconceptions more than the "I don't know" answer.

The aim is not to compare these grading results, but to try and evaluate how the scores change from pretest to posttest when using different gradings: the lenient gradings will show whether basic understanding was achieved while the strict grading will show whether misconceptions were introduced. Similarly, I used Sankey diagrams to show the flows between groups of participants having similar overall grades on question groups.

grading	severe misconception	mild misconception	does not know	correct
base	0	0	0	1
lenient	0	2	1	3
category-lenient	0	2	1	2
strict	0	1	2	3

Table 6.5: Gradings: points per answer type

**Basic understanding** In this paragraph I define *basic understanding* as fulfilled when participants choose answers that are of the same nature as the correct answer. If the correct answer to a question is a BioBrick, the participant has reached basic understanding if they choose *any* BioBrick instead of a Device or another game element listed in the possible answers. Figure 6.22 shows that the objective of basic understanding as I defined it is not clearly reached for the BioBrick questions.

The scores on the left and on the right are numbers between 0 and 10, corresponding to the sorted range of possible scores as there are five questions related to

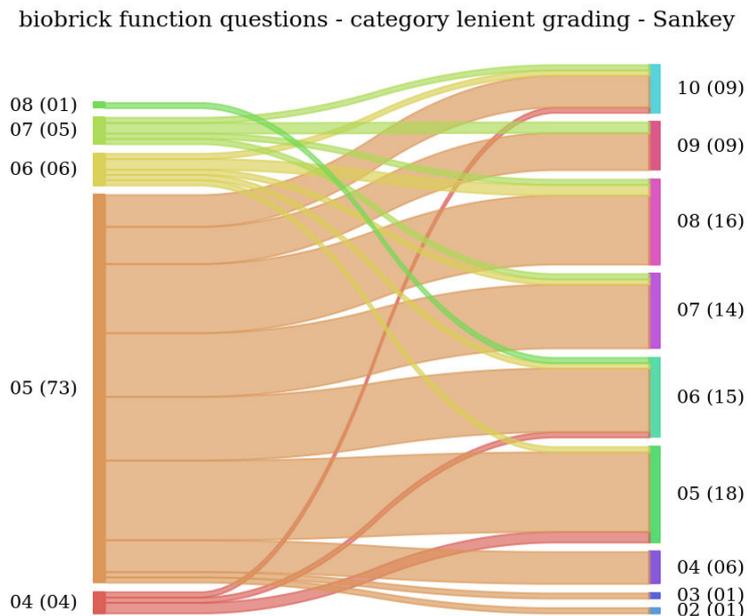


Figure 6.22: Sankey diagram of scores on BioBrick function questions using a category-lenient grading

BioBricks, two points per answer of correct nature, one point per "I don't know" and zero point per answer of incorrect nature. The scores are sorted vertically from the lowest on the bottom to the highest on the top. On the left (respectively right) are the pretest (respectively posttest) scores, and in parentheses are the number of participants who got these scores. The colored strips represent the participants who got different pretest and posttest scores, the widths of the strips being proportional to the number of participants who did so. The colors of these strips depend on the pretest scores of those participants. Scores below five denote at least one answer of incorrect nature (7 participants out of 89), and scores equal to ten denote perfect understanding of the nature of the answers (9 participants out of 89). All scores in between show at least one misunderstanding on the nature of a proposed answer.

This cannot be explained simply by the vocabulary problem already discussed in section 6.1.4 and section 6.2.2. Here, the questions are asked by referring to the visual representation of the BioBricks in the game, rather than to their names. Maybe the description of the BioBrick functions themselves use a vocabulary that is not mastered by the participants. To assert this hypothesis, a further study may employ a visual way to describe the function of the BioBrick instead of a sentence.

**Misconceptions** In this paragraph I will focus on participants who answered correctly or "I don't know" in the pretest and who answered incorrectly in the

posttest. To detect such participants I will use the *strict* grading defined in table 6.5 and plot Sankey diagrams. As a first and simple example, I will only use the question related to genotype and phenotype: "In order to modify the abilities of the bacterium, you have to..." (question 11 in the annex C.4): see figure 6.23.

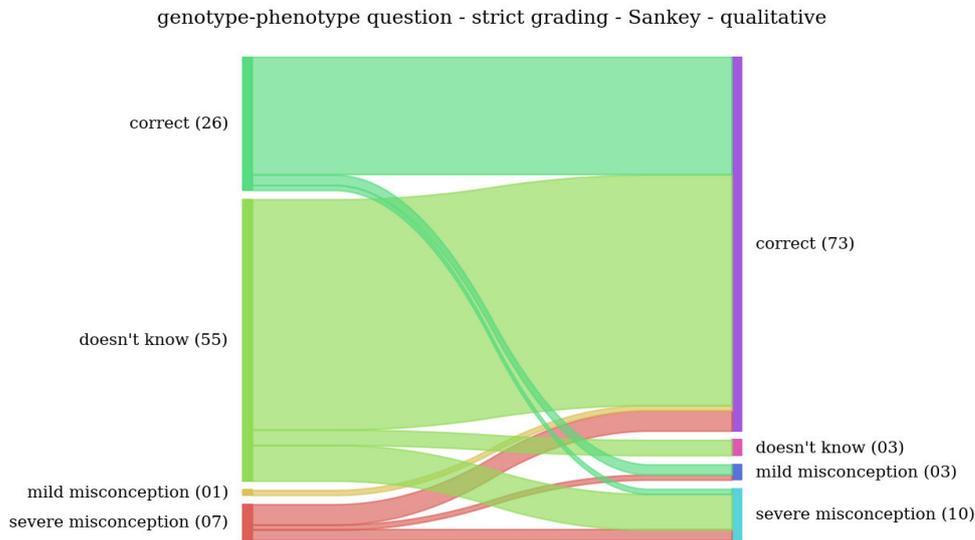


Figure 6.23: Sankey diagram of answers on the genotype-phenotype question using a strict grading

The graph shows that mild and severe and mild misconceptions are introduced to 8 and 2 participants respectively, to put into perspective with the 50 participants who corrected their answer.

As a second example, I will also use a single question related to inducible promoters: "Last question. Next page only contains remarks. Guess: you have crafted a functional device containing an arabinose-induced promoter and an arabinose Coding Sequence (CDS). What will happen?" (question 18 in the annex C.4): see figure 6.24.

The graph shows that mild and severe misconceptions are introduced to 21 and 4 participants respectively, to put into perspective with the 13 participants who corrected their answer. This question is the hardest to read, to understand, and to answer (it refers to the latter sections of the game), and is the very last one of the questionnaire, which may explain why so many participants still answered "I don't know" after playing.

Lastly I will analyze a group of questions related to inducible promoters: question 15 - device 3, question 15 - device 4, and question 18, in the annex C.4. Figure 6.25 leads us to the same conclusion: even when considering these three questions together, most of the participants end up with a decreased posttest score, showing that misconceptions were introduced.

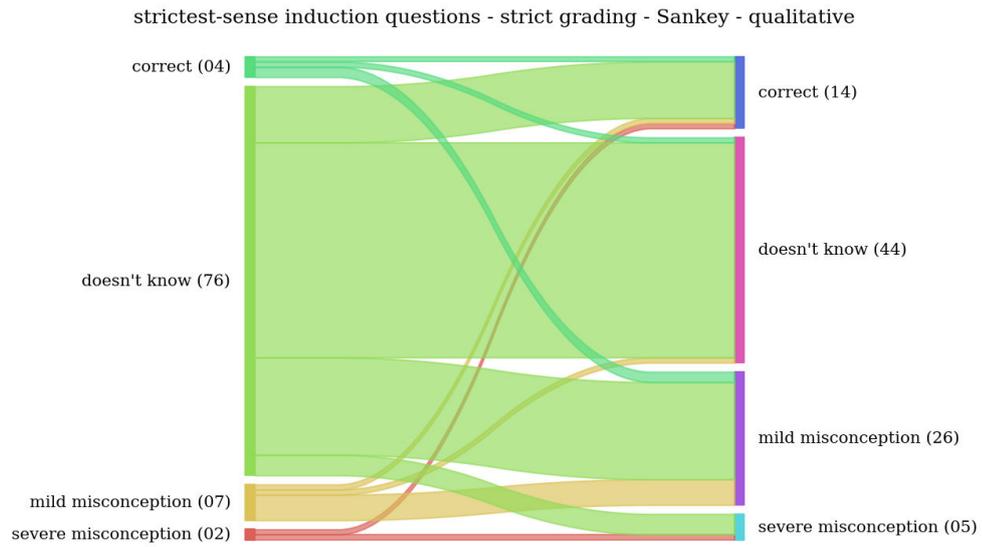


Figure 6.24: Sankey diagram of answers on one induction question using a strict grading

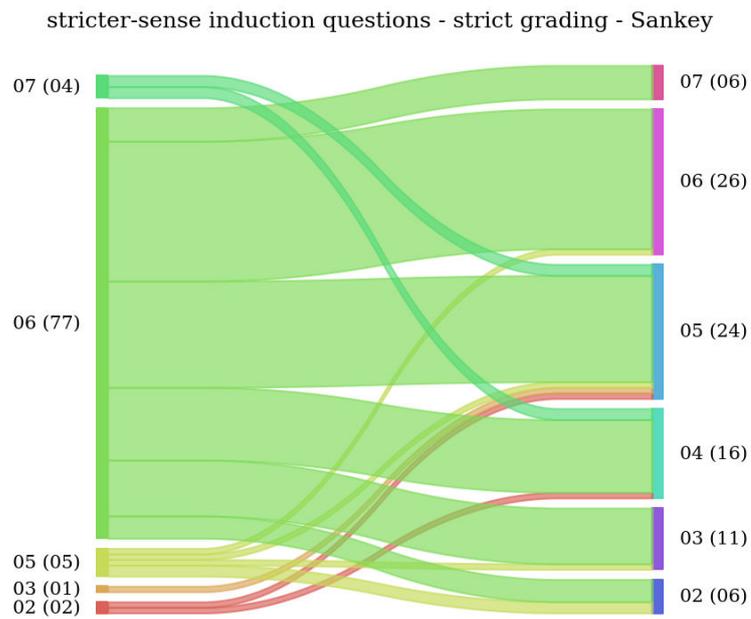


Figure 6.25: Sankey diagram of answers on three induction questions using a strict grading

The same kind of analysis on the 8 device questions shows that the game induced misconceptions about BioBrick assembly in 17 participants.

These results make it possible to further discuss our starting hypothesis: textbfhypothesis H2b: "the use of a game can make citizens understand basic notions about synthetic biology".

- a) "the simplified link between genotype and phenotype": only a fraction of participants end up having a misconception regarding this notion.
- c) "the BioBrick simplified grammar (Promoter - RBS - Coding Sequence - Terminator)": misconceptions are introduced by the game.
- d) "the names and functions of the bricks (condition - quantity - function - end)": misconceptions are introduced by the game.
- e) "the advanced notion of inducible promoters": misconceptions are introduced by the game.

This ultimately makes it possible to also confirm the hypothesis **H2c: "the use of a game can introduce misconceptions in the minds of citizens"**.

**Change in interest** The same type of visualization can be applied to the questions pertaining to interest, in order to more finely analyze how the answers changed from the pretest to the posttest. These questions are the first question of the questionnaire feature in annex C.4: "Are you interested in learning more about..." with the four sub-questions "Biology", "Synthetic biology", "Video games", and "Engineering". For this part of the study I chose to use the sample of 126 participants whose pretest and posttest data are exploitable rather than the sample of 89 participants who additionally answered optional questions. The ratings on the left and on the right are numbers between 1 and 5, corresponding to the sorted range of possible answers from "Not at all" = 1 to "Extremely" = 5. The ratings are sorted vertically from the lowest on the bottom to the highest on the top. On the left (respectively right) are the pretest (respectively posttest) ratings, and in parentheses are the number of participants who self-assessed their interest to each rating. The colored strips represent the participants who changed their interest rating between pretest and posttest, the widths of these strips being proportional to the number of participants who did so. The colors of these strips depend on the pretest rating of those participants.

The four graphs are shown in figure 6.26, figure 6.27, figure 6.28, and figure 6.29.

These graphs show a phenomenon of polarization of opinions. As was shown in section 6.2.2, there is no clear change of opinion from positive to negative, but a slight trend of decrease in interest: this finer analysis shows that for instance Engineering sees negative opinions rise from 26 to 36 (20.8% to 28.8%) and positive opinions decrease from 65 to 56 (52.0% to 44.8%) while neutral opinions are stable (27.2% to 26.4%).

"Want to learn more about Biology" - Sankey 1-5 score variation

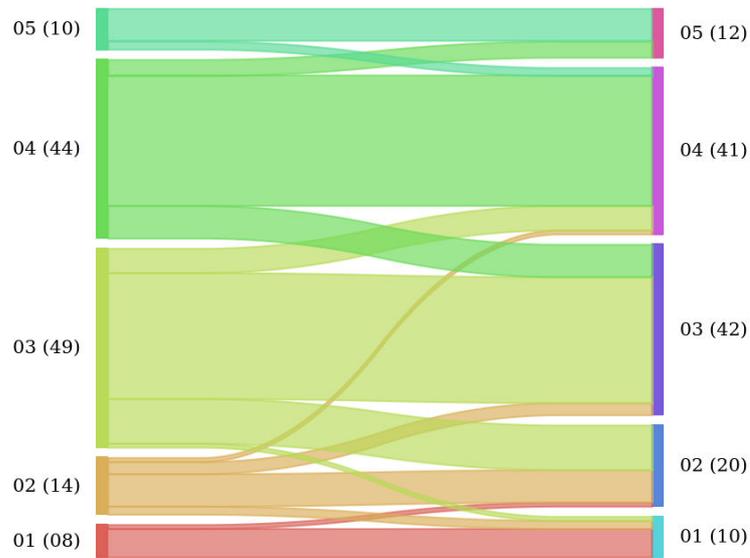


Figure 6.26: Change in interest in Biology

"Want to learn more about Synthetic biology" - Sankey 1-5 score variation

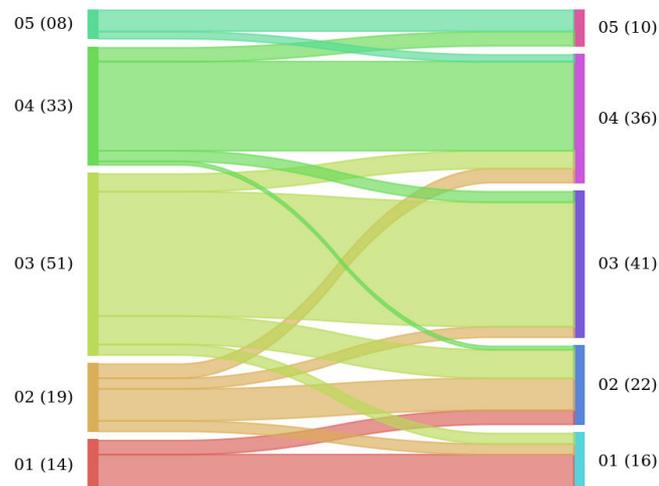


Figure 6.27: Change in interest in Synthetic Biology

Another trend is that extreme opinions increase in size while neutral opinions decrease. This polarization is still variable from theme to theme. All of the four themes of interest see their pool of neutral opinions decrease, from 42 to 30

"Want to learn more about Video games" - Sankey 1-5 score variation

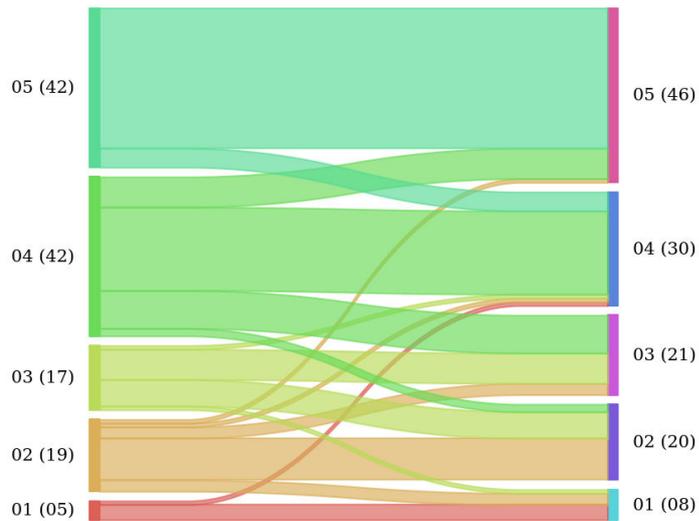


Figure 6.28: Change in interest in Video Games

"Want to learn more about Engineering" - Sankey 1-5 score variation

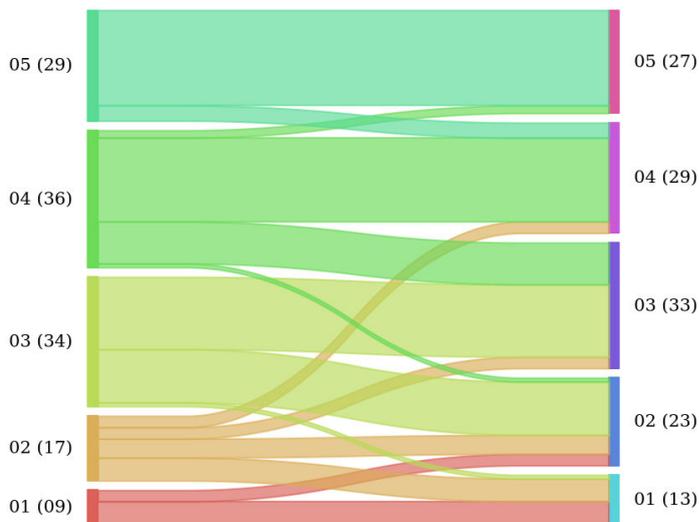


Figure 6.29: Change in interest in Engineering

(Video Games, 33.6% to 24.0%) to 34 to 33 (Engineering, 27.2% to 26.4%), with proportionally more ratings in the extremes in both negative and positive opinions. For instance for Engineering, extreme negative opinions rise from 34.6% to 36.1%

of negative opinions and extreme positive opinions rise from 44.6% to 48.2% of opinions. It shows that participants are slightly less interested and their opinions are slightly more polarized.

opinion type	extreme negative opinions		extreme positive opinions	
	pretest	posttest	pretest	posttest
Biology	36.4%	33.3%	18.5%	22.6%
Synthetic Biology	42.4%	42.1%	19.5%	21.7%
Video Games	20.8%	28.6%	50.0%	60.5%
Engineering	34.6%	36.1%	44.6%	48.2%

Table 6.6: Proportion of extreme opinions among positive and negative opinions

### 6.2.3 Surveys and game metrics: data mining

The exploratory analysis demonstrated correlations between the time spent playing and survey answers. We also wanted to assert the relationship between other events in the game and the outcomes of the survey, in order to evaluate the possibility of using game metrics instead of questionnaires to assess learning outcomes. We used data mining techniques to analyze the surveys and the tracking data.

#### Predictability of the scores using game metrics

The overall posttest score cannot be predicted from tracking data using machine learning prediction: a maximum accuracy of 0.39 was reached, with a high standard deviation of 0.26. The scores on individual questions can be predicted with better accuracy, with a mean accuracy of 0.71. This answers research question RQ7 *Quiz-based assessment and automated tracking: How comparable are learning metrics computed from questionnaires and from automated remote tracking data? Can quiz-based assessment be replaced by automated tracking?* and its associated hypothesis *H6b: "The learning outcomes are predictable using game metrics."*

These accuracies were computed using classifiers and cross-validation of scores in Python with the scikit-learn package. First, an automated optimization using linear regression reached a prediction accuracy of 0.36 with a high standard deviation of 0.31. This automated process had to optimize through all the features (85) from the tracking data and therefore was likely not to be optimal. That is why in a second time I identified the most weighted features in the linear regression (table 6.7) and used a brute-force algorithm to search for a set of features including these and a reduced number of features to optimize a polynomial regression prediction and obtain a mean accuracy of 0.39.

Feature code	Weight
ch06completion	1.669966e-01
ch02completion	8.394467e-01
ch07total	9.338385e-01
ch05total	2.005835e+00

Table 6.7: Weight of features in automated regression of the posttest score from tracking data

**Separate questions** The method used to compute the accuracy of prediction of question score is the same. Using only a subset of tracking data, the list of completion times, the accuracy of predictions has a mean of 0.71 and can go up to 0.96 (in a probable case of overfitting). Table 6.8 lists the computed accuracies.

Using the whole set of tracking data, the results are comparable: the accuracy of predictions has a mean of 0.71 too (table 6.9). This shows that completion times are the best predictors for the scores on individual questions.

Question code	Mean accuracy	Accuracy variance
Genotype and phenotype	0.797695	0.055617
BioBricks and devices composition	0.630719	0.126737
Ampicillin antibiotic	0.640764	0.073011
Name: Plasmid	0.597386	0.090739
Function: TER	0.698452	0.056737
Name: PR	0.684623	0.032357
Function - game: CDS	0.822910	0.055844
Name: TER	0.525593	0.127207
Function - biology: CDS	0.684314	0.063287
Name: RBS	0.854489	0.057721
Example: CDS	0.663158	0.092246
Name: CDS	0.764052	0.021519
Function: PR	0.775163	0.035523
Function: RBS	0.594771	0.148511
Function: Plasmid	0.706536	0.137028
Name: Operator XXX	0.955556	0.022222
Device: RBS:PCONS:FLHDC:TER XXX	0.809529	0.043103
Device: PCONS:RBS:FLHDC:TER	0.696732	0.082529
Device: PBAD:RBS:GFP:TER	0.784830	0.068963
Device: PBAD:GFP:RBS:TER XXX	0.664706	0.082275
Device: GFP:RBS:PCONS:TER XXX	0.684107	0.097740
Device: PCONS:GFP:RBS:TER XXX	0.650327	0.101444
Device: AMPR:RBS:PCONS:TER XXX	0.753595	0.107528
Device: RBS:PCONS:AMPR:TER XXX	0.684967	0.031250
Green fluorescence	0.604747	0.073020
Unequip the movement device: effect	0.538562	0.048884
Device: PBAD:RBS:ARA:TER	0.820915	0.052873

Table 6.8: Mean and variance of prediction accuracy on different questions based on completion times

Similarly to the previous section, the overall posttest score cannot be predicted from tracking data and pretest surveys: using machine learning prediction, a maximum accuracy of 0.44 was reached, with a high standard deviation of 0.26. Searching on a greater number of features could have yielded a better optimum.

### 6.2.4 Comparison of phase 1 and phase 2

In order to evaluate the difference between the two versions of the game I compared the scores obtained by the 89 participants in phase 1 and the 30 participants in phase 2 of the experiment (as presented in section 6.1.1). I also compared the times spent in the game, expecting that phase 2 participants would spend less time in the earlier sections of the game.

Question code	Mean accuracy	Accuracy variance
Question code	Mean	Var
Genotype and phenotype	0.740832	0.049212
BioBricks and devices composition	0.607190	0.097480
Ampicillin antibiotic	0.637702	0.128454
Name: Plasmid	0.460784	0.022068
Function: TER	0.639628	0.030997
Name: PR	0.697626	0.050671
Function - game: CDS	0.799381	0.068934
Name: TER	0.571999	0.062974
Function - biology: CDS	0.730719	0.039030
Name: RBS	0.845201	0.045501
Example: CDS	0.663158	0.027511
Name: CDS	0.709150	0.070485
Function: PR	0.786275	0.024068
Function: RBS	0.650327	0.107357
Function: Plasmid	0.685621	0.095956
Name: Operator XXX	0.955556	0.022222
Device: RBS:PCONS:FLHDC:TER XXX	0.759993	0.106094
Device: PCONS:RBS:FLHDC:TER	0.675163	0.078557
Device: PBAD:RBS:GFP:TER	0.776711	0.053103
Device: PBAD:GFP:RBS:TER XXX	0.629412	0.054863
Device: GFP:RBS:PCONS:TER XXX	0.696457	0.091441
Device: PCONS:GFP:RBS:TER XXX	0.684967	0.047022
Device: AMPR:RBS:PCONS:TER XXX	0.729412	0.092117
Device: RBS:PCONS:AMPR:TER XXX	0.810458	0.074205
Green fluorescence	0.685208	0.026824
Unequip the movement device: effect	0.630065	0.071699
Device: PBAD:RBS:ARA:TER	0.876471	0.041306

Table 6.9: Mean and variance of prediction accuracy on different questions based on tracking data

When comparing the scores obtained, the t-value was  $t=0.93$  and the p-value  $p=0.36$ , failing to reject the null hypothesis that both game versions are equivalent. When comparing the times spent, the p-values were always above 0.05. The completion rate of the phase 1 and phase 2 participants was 30.3% and 13.3% respectively. These results may be explained by a new flaw in game design introduced in this new version, by the relatively small number of participants in phase 2 (30), prone to statistical insignificance and to sampling bias, especially in a museum where the type of visitors varies along the year and depends on the days of the week.

### 6.2.5 Limitations

The cohort of participants was too small to ascertain the effects of the game on the subsamples, especially on the second phase of the experiment ( $n = 30$ ).

Additionally, the experiment was not designed to measure how long the learning effect carried on on the participants. Further research should be done to know whether this knowledge persists, the problem being that it would necessitate many experimental subjects. In a perfectly rigorous protocol, to measure how much participants remember after the durations  $\Delta t_1$ ,  $\Delta t_2$ , ..., and  $\Delta t_n$ ,  $n$  cohorts of sufficient size have to be recruited. Using the same cohort each time would not work: the evaluation method used here, multiple-choice questionnaires, induces learning (Valerie J. Shute, Hansen, and Almond, 2008; Kleij et al., 2012). Consequently the results would be similar to the Forgetting Curve (Bailey, 1989; Averell and Heathcote, 2011).

# Chapter 7

## Conclusions and prospectives

### 7.1 Conclusions

This section summarizes the contributions this thesis makes by answering the research questions and hypotheses formulated in section 3.

#### 7.1.1 Research questions

##### **RQ1: Academic education**

This experiment did not include in-school formal education. Previous in-class experiments (section 5.3) yielded results tending to show that students learned synthetic biology content with the game, with misconceptions, and were engaged in the activity. More data are needed to confirm this trend and to assess more precisely which notions are mastered by the students.

##### **RQ2: Popularization and lifelong learning**

**H2a: the use of a game can increase the motivation and curiosity in discovering synthetic biology.** In this experiment, curiosity was stable (section 6.2.2), but more importantly, the cohort was eventually slightly more polarized (section 6.2.2). This cannot confirm the hypothesis. However the game itself was enjoyed by the participants (6.2.2), figure 6.20.

**H2b: the use of a game can make citizens understand basic notions about synthetic biology:** a) the simplified link between genotype and phenotype was shown to be understood in section 6.2.2, figure 6.20. and section 6.2.2. The score improved in a statistically significant manner, event with a strict grading policy.

b) The fact that BioBricks and devices are DNA sequences was shown to be understood in section 6.2.2, figure 6.20.

c) the BioBrick simplified grammar (Promoter - RBS - Coding Sequence - Terminator) was shown to be understood in section 6.2.2, figure 6.20. The score

improved globally on the 8 questions on correct BioBrick assembly, in a statistically significant manner, even with a strict grading policy.

d) section 6.2.2 shows that the names of the bricks could not be shown as learned, but the simplified role of each kind of brick (condition - quantity - function - end) could in part, with introduced misconceptions (section 6.2.2).

e) the advanced notion of inducible promoters could not be shown as mastered by the participants. Looking at the overall results on inducible promoters (section 6.2.2), there is a slight improvement, statistically significant, but not with a strict grading policy, in which case a strong decrease is observed, revealing that misconceptions were introduced (section 6.2.2).

**H2c: the use of a game can introduce misconceptions in the minds of citizens.** In addition to the early experiments (see section 6.2.2), this hypothesis was confirmed in section 6.2.2 as described in the previous paragraph.

**RQ3: Learning efficiency, motivation and player characteristics**

**H3a: age correlates negatively with score; gender does not.** No correlation was found in section 6.1.3.

**H3b: age correlates negatively with motivation; gender does not.** No correlation was found in section 6.1.3 between age and gender and motivation.

**H3c: scores correlate positively with interest and education in biology.** This hypothesis is dealt with in section 6.1.3. Interest in biology is weakly correlated to score and score increase. Curiosity and education in biology are correlated to score and score increase.

**H3d: motivation correlates positively with interest and education in biology.** Interest and curiosity in biology correlate positively with motivation, education in biology does not (section 6.1.3).

**H3e: scores correlate positively with interest and practice in games.** No correlation was found between scores and interest, curiosity, and practice in video games (section 6.1.3).

**H3f: motivation correlates positively with interest and practice in games.** Interest and practice in video games correlate positively with motivation, curiosity does not (section 6.1.3).

**RQ4: Playing duration and player characteristics**

**H4a: age correlates positively with playing duration to complete the game; gender does not.** No global correlation was found in section 6.1.4 between age and gender and playing durations. In the very first sections, completion times and spent times correlate positively with age and being female.

**H4b: playing duration to complete the game correlates negatively with interest and education in biology.** No overall correlation was found in section 6.1.4 between interest in biology and total playing duration. However, the total completion time does correlate negatively with interest in biology.

**H4c: playing duration to complete the game correlates negatively with interest and practice in games.** No overall correlation was found in section 6.1.4 between total playing durations and interest in games and gaming practice. In the very first sections, completion times and spent times correlate negatively with interest in games and gaming practice, and the trend is opposite for time spent in later sections.

**RQ5: Player characteristics and implicit, explicit content**

This research question could not be evaluated due to constraints in the implementation of the experiment. Preliminary in-class tests tended to show that there was no difference in version 1.12 (section 5.3).

**RQ6: Quiz-based assessment and automated tracking**

**H6a: In a linear game where each puzzle is compulsory to solve one after the other to finish the game, reaching some thresholds can be equivalent to validating the assimilation of a notion.** Certain questions indicate checkpoint 5 as a learning threshold (section 6.2.1).

**H6b: The learning outcomes are predictable using game metrics.** The overall posttest score cannot be predicted from tracking data using machine learning prediction, a maximum accuracy of 0.39 is reached (section 6.2.3). Higher accuracies (mean: 0.71) are reached for individual questions. Taking into account the pretest, the accuracy of the prediction of the posttest score reaches 0.44.

**RQ7: Threshold effect**

**H7a: After reaching a certain point in the game, most of the learning is achieved: going further will not increase the learning outcomes.** There is no overall threshold effect in the game (section 6.2.1).

These conclusions, when mixed with feedbacks by the participants, can be reformulated in terms of usefulness, usability and acceptability.

### 7.1.2 Usefulness, Usability, Acceptability

#### Usefulness

In terms of learning, the game has a measurable immediate impact on knowledge in SB.

No difference was found across self-declared genders. The older the participants were, the less the learning outcomes were.

Further research is needed to ascertain medium-term and long-term effects on learning and motivation.

Some correlations were established between actions in the game and results in the assessments, but did not lead to a precise prediction of the score.

#### Usability

The experiment was not designed to properly measure usability, but a qualitative estimate can be set forth: without explanations, the participants were able to move the bacterium, click on buttons, craft, and discover by themselves what they had to do. Only a minority reported having trouble understanding what the goal of the game was. However, an important part of the players were stuck in two bottleneck chapters of the game.

Another important usability problem was related to the assessment process: participants were confused about the pretest-posttest process. It can be speculated that this is due to insufficient explanations and presentation problems from the experimenters' side, and unfamiliarity to experimental protocols for the participants.

#### Acceptability

The experiment was not designed to properly measure acceptability. The proxy measurements - self-assessed engagement and rate of leaving - are only indicative. The game reached good acceptability, because (1) the game was enjoyed by 80% of the participants (figure 6.20) (2) only a minority of volunteers left before the end (67 at most out of 193, from figure 6.2). Of course there is a selection bias as only *volunteers* participated in the experiment. Acceptability would be best measured with a cohort of citizens encouraged to play with an external reward such as money. Acceptability was estimated on previous versions of the game with students having a compulsory session of Hero.Coli and was estimated to 90% of students having fun playing the game.

## 7.2 Prospectives

Hero.Coli is an open-source video game and is continuously being improved in its implementation and evaluation. As this thesis demonstrated, there are still many

possible improvements. Improvements to address a bigger and more various audience: new platforms could be reached (Android, iPhone), new game modes can be developed and explored that focus on competition (multiplayer or challenge levels). Engagement can be increased through an improved and expanded storyline, further uses of UX principles, and improved, more progressive and adaptive tutorials to help players understand the game mechanics. Tutorials have indeed proven critical in this thesis with the differences induced between the first versions of the game - deprived of proper tutorials - and the last ones with adaptive tutorials. The problem of the teaching of induced promoters has yet to be solved, in addition to inhibited promoters, not yet taught in the game. This would also contribute to better learning, in addition to new content (BioBricks), new mechanics (more categories of BioBricks and a laxer grammar of BioBrick combination), improvements on the simulator (if not an integration of the whole-cell model, then at least the integration of some pathways, maybe in a separate game mode), as well as more transmedia integration (for instance, specific single levels that illustrate MOOCs, or videos integrated in the game). Additionally, improvements can be made to boost the research interest of the project: citizen science can still be integrated through implemented challenges from synthetic biology research. Research on the effectiveness of digital game-based learning can also be furthered by applying results from Trace Theory that were not applied in this thesis, using Petri networks and Markov chains (É. Sanchez, Ney, and Labat, 2011), or already implemented solutions similar to Undertracks (Bouhineau et al., 2013). This would help analyze and identify user behavior issues relative to synthetic biology practice. Data gathering can be improved through the use of incentives: answering the posttest could be rewarded by the unlocking of content such as new characters. *E. coli* is the default avatar in the game but the well-studied *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, or *Saccharomyces cerevisiae* (Macklin, Ruggero, and Markus W Covert, 2014) could be additional characters.

Beyond game and protocol improvements, the question of integrating the game itself in a learning scenario is still open. Digital game-based learning proves more effective when prepared with, accompanied by, and discussed with the teacher. Conducting controlled experiments on several classes of students would make it possible to compare the use of a game in synthetic biology teaching to the classic approaches. Long-term studies would also be of interest to assess the retention induced by a synthetic biology game, especially on aspects which were learned with different successes: vocabulary, icons, and mechanics. This would help teachers create or adapt their learning scenarios to complement digital game-based learning of synthetic biology.



# Bibliography

- Adams, Bryn L. (Dec. 16, 2016). “The Next Generation of Synthetic Biology Chassis: Moving Synthetic Biology from the Laboratory to the Field”. In: *ACS Synthetic Biology* 5(12), pp. 1328–1330. ISSN: 2161-5063, 2161-5063. DOI: 10.1021/acssynbio.6b00256.
- Ahn, L. von (June 2006). “Games with a purpose”. In: *Computer* 39(6), pp. 92–94. ISSN: 0018-9162. DOI: 10.1109/MC.2006.196.
- Amiel, Tel and Thomas C. Reeves (Oct. 2008). “Design-Based Research and Educational Technology: Rethinking Technology and the Research Agenda”. In: *Educational Technology & Society* 11(4), p. 12.
- Amy Bruckman (Mar. 1999). “Can Educational Be Fun?” In: Game Developer’s Conference. San Jose, California.
- Andrianantoandro, Ernesto et al. (May 16, 2006). “Synthetic biology: new engineering rules for an emerging discipline”. In: *Molecular Systems Biology* 2. ISSN: 1744-4292. DOI: 10.1038/msb4100073.
- Annetta, Leonard A. et al. (Aug. 2009). “Investigating the impact of video games on high school students’ engagement and learning about genetics”. In: *Computers & Education* 53(1), pp. 74–85. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2008.12.020.
- Apperley, Thomas H. (Mar. 2006). “Genre and game studies: Toward a critical approach to video game genres”. In: *Simulation & Gaming* 37(1), pp. 6–23. ISSN: 1046-8781, 1552-826X. DOI: 10.1177/1046878105282278.
- Ask, Kristine (2017). “The Value of Calculations: The Coproduction of Theorcraft and Player Practices”. In: 36. ISSN: 0270-4676. DOI: 10.1177/0270467617690058.
- Aucamp, Janine et al. (Dec. 1, 2016). “A historical and evolutionary perspective on the biological significance of circulating DNA and extracellular vesicles”. In: *Cellular and Molecular Life Sciences* 73(23), pp. 4355–4381. ISSN: 1420-682X, 1420-9071. DOI: 10.1007/s00018-016-2370-3.
- Avellaneda, Rafael Pardo and Kristin Hagen (2016). “Synthetic biology: public perceptions of an emergent field”. In: *Synthetic Biology Analysed*. Springer, pp. 127–170.
- Averell, Lee and Andrew Heathcote (Feb. 2011). “The form of the forgetting curve and the fate of memories”. In: *Journal of Mathematical Psychology* 55(1), pp. 25–35. ISSN: 00222496. DOI: 10.1016/j.jmp.2010.08.009.

- Bailey, Charles D. (Mar. 1989). “Forgetting and the Learning Curve: A Laboratory Study”. In: *Management Science* 35(3), pp. 340–352. ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.35.3.340.
- Bamparopoulos, Giorgos et al. (Feb. 9, 2016). “Towards exergaming commons: composing the exergame ontology for publishing open game data”. In: *Journal of Biomedical Semantics* 7, p. 4. ISSN: 2041-1480. DOI: 10.1186/s13326-016-0046-4.
- Barone, Jonathan et al. (2015). “Nanocrafter: Design and Evaluation of a DNA Nanotechnology Game.” In: *FDG*.
- Bitzer, D., P. Braunfeld, and W. Lichtenberger (Dec. 1961). “PLATO: An Automatic Teaching Device”. In: *IRE Transactions on Education* 4(4), pp. 157–161. ISSN: 0893-7141. DOI: 10.1109/TE.1961.4322215.
- Blunch, Niels J. (May 1984). “Position Bias in Multiple-Choice Questions”. In: *Journal of Marketing Research (JMR)* 21(2), pp. 216–220. ISSN: 00222437.
- Bohannon, John (Oct. 24, 2008). “Flunking Spore”. In: *Science* 322(5901), pp. 531–531. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.322.5901.531b.
- Bouhineau, Denis et al. (2013). “Conception et mise en place d’un entrepôt de traces et processus de traitement EIAH: UnderTracks”. In: *EIAH 2013-6e Conférence sur les Environnements Informatiques pour l’Apprentissage Humain*. IRIT Press 2013, pp. 41–42.
- Bouvier, P. et al. (July 2013). “Using Traces to Qualify Learner’s Engagement in Game-Based Learning”. In: *2013 IEEE 13th International Conference on Advanced Learning Technologies*. 2013 IEEE 13th International Conference on Advanced Learning Technologies, pp. 432–436. DOI: 10.1109/ICALT.2013.132.
- Bouvier, Patrice, Elise Lavoué, et al. (May 2013). “Identifying Learner’s Engagement in Learning Games: a Qualitative Approach based on Learner’s Traces of Interaction”. In: *5th International Conference on Computer Supported Education (CSEDU 2013)*. Aachen, Germany, pp. 339–350.
- Bouvier, Patrice, Karim Sehaba, and Elise Lavoué (2014). “A trace-based approach to identifying users’ engagement and qualifying their engaged-behaviours in interactive systems: application to a social game”. In: *User Modeling and User-Adapted Interaction* 24(5), pp. 413–451. DOI: 10.1007/s11257-014-9150-2.
- Boyle, Elizabeth A. et al. (Mar. 2016). “An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games”. In: *Computers & Education* 94, pp. 178–192. ISSN: 03601315. DOI: 10.1016/j.compedu.2015.11.003.
- Boyle, Elizabeth, Thomas M. Connolly, and Thomas Hainey (Jan. 1, 2011). “The role of psychology in understanding the impact of computer games”. In: *Entertainment Computing. Serious Games Development and Applications* 2(2), pp. 69–74. ISSN: 1875-9521. DOI: 10.1016/j.entcom.2010.12.002.
- Bügl, Hans et al. (June 2007). “DNA synthesis and biological security”. In: *Nature Biotechnology* 25(6), pp. 627–629. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt0607-627.

- Caillois, Roger (1958). “Les Jeux et les Hommes: Le masque et le vertige (Man, Play and Games)”. In: *Paris: Gallimard*.
- Caillois, Roger and Elaine P. Halperin (Sept. 1955). “The Structure and Classification of Games”. In: *Diogenes* 3(12), pp. 62–75. ISSN: 0392-1921, 1467-7695. DOI: 10.1177/039219215500301204.
- Cameron, Brian and Francis Dwyer (July 2005). “The Effect of Online Gaming, Cognition and Feedback Type in Facilitating Delayed Achievement of Different Learning Objectives”. In: *Journal of Interactive Learning Research* 16(3), pp. 243–258. ISSN: 1093-023X.
- Carl, Jim (2009). “Industrialization and Public Education: Social Cohesion and Social Stratification”. In: *International Handbook of Comparative Education*. Springer International Handbooks of Education. Springer, Dordrecht, pp. 503–518. ISBN: 978-1-4020-6402-9 978-1-4020-6403-6. DOI: 10.1007/978-1-4020-6403-6\_32.
- Carter, Marcus and Martin Gibbs (2013). “eSports in EVE Online: Skullduggery, Fair Play and Acceptability in an Unbounded Competition”. In: p. 8.
- Caterina, Michael J. et al. (Oct. 1997). “The capsaicin receptor: a heat-activated ion channel in the pain pathway”. In: *Nature* 389(6653), pp. 816–824. ISSN: 1476-4687. DOI: 10.1038/39807.
- Church, George M., Michael B. Elowitz, et al. (Apr. 2014). “Realizing the potential of synthetic biology”. In: *Nature Reviews Molecular Cell Biology* 15(4), pp. 289–294. ISSN: 1471-0072. DOI: 10.1038/nrm3767.
- Church, George M. and Edward Regis (2012). *Regenesis: how synthetic biology will reinvent nature and ourselves*. Basic Books: New York, NY. 284 pp. ISBN: 978-0-465-02175-8 978-0-465-03329-4.
- Cira, Nate J. et al. (Mar. 25, 2015). “A Biotic Game Design Project for Integrated Life Science and Engineering Education”. In: *PLOS Biology* 13(3), e1002110. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1002110.
- Clauzel, Damien, Karim Sehaba, and Yannick Prié (Jan. 2011). “Enhancing synchronous collaboration by using interactive visualisation of modelled traces”. In: *Simulation Modelling Practice and Theory*. Modeling and Performance Analysis of Networking and Collaborative Systems 19(1), pp. 84–97. ISSN: 1569190X. DOI: 10.1016/j.simpat.2010.06.021.
- Conati, Cristina and Micheline Manske (2009). “Evaluating Adaptive Feedback in an Educational Computer Game”. In: *Intelligent Virtual Agents*. Ed. by Zsófia Ruttkay et al. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 146–158. ISBN: 978-3-642-04380-2.
- Connolly, Thomas M. et al. (Sept. 2012). “A systematic literature review of empirical evidence on computer games and serious games”. In: *Computers & Education* 59(2). WOS:000305036400043, pp. 661–686. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2012.03.004.
- Cooper, Seth et al. (Aug. 2010). “Predicting protein structures with a multiplayer online game”. In: *Nature* 466(7307), pp. 756–760. ISSN: 1476-4687. DOI: 10.1038/nature09304.

- Covert, Markus W. et al. (Sept. 15, 2008). “Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*”. In: *Bioinformatics* 24(18), pp. 2044–2050. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btn352.
- Csikszentmihalyi, Mihaly (1997). “Flow and the psychology of discovery and invention”. In: *HarperPerennial, New York* 39.
- Davidovitch, Lior, Avi Parush, and Avy Shtub (Apr. 1, 2008). “Simulation-based learning: The learning–forgetting–relearning process and impact of learning history”. In: *Computers & Education* 50(3), pp. 866–880. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2006.09.003.
- Davison, John (Feb. 1, 2010). “GM plants: Science, politics and EC regulations”. In: *Plant Science* 178(2), pp. 94–98. ISSN: 0168-9452. DOI: 10.1016/j.plantsci.2009.12.005.
- Dede, Chris (2009). “Comparing Frameworks for “21st Century Skills””. In: *21st Century Skills: Rethinking How Students Learn*, p. 16.
- Dekel, Erez and Uri Alon (July 2005). “Optimality and evolutionary tuning of the expression level of a protein”. In: *Nature* 436(7050), pp. 588–592. ISSN: 1476-4687. DOI: 10.1038/nature03842.
- Dimitrov, Dimiter M. and Jr Rumrill (Jan. 1, 2003). “Pretest-posttest designs and measurement of change”. In: *Work* 20(2), pp. 159–165. ISSN: 1051-9815.
- Djaouti, Damien, Julian Alvarez, and Jean-Pierre Jessel (2011). “Classifying serious games: the G/P/S model”. In: *Handbook of research on improving learning and motivation through educational games: Multidisciplinary approaches*. IGI Global, pp. 118–136.
- Djaouti, Damien, Julian Alvarez, Jean-pierre Jessel, et al. (2008). *Play, Game, World: Anatomy of a Videogame*.
- Ducheneaut, Nicolas et al. (2007). “The Life and Death of Online Gaming Communities: A Look at Guilds in World of Warcraft”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '07. ACM: New York, NY, USA, pp. 839–848. ISBN: 978-1-59593-593-9. DOI: 10.1145/1240624.1240750.
- Dunlap, Peter and JL Pecore (2009). “The Effects of Gaming on Middle and High School Biology Students’ Content Knowledge and Attitudes toward Science”. In: *Studies in Teaching*, pp. 19–36.
- Ebbinghaus, Hermann (1885). *Über das gedächtnis: untersuchungen zur experimentellen psychologie*. Duncker & Humblot.
- Elowitz, Michael B. and Stanislas Leibler (Jan. 2000). “A synthetic oscillatory network of transcriptional regulators”. In: *Nature* 403(6767), pp. 335–338. ISSN: 1476-4687. DOI: 10.1038/35002125.
- Elowitz, Michael B., Arnold J. Levine, et al. (Aug. 16, 2002). “Stochastic Gene Expression in a Single Cell”. In: *Science* 297(5584), pp. 1183–1186. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1070919.
- Endy, Drew (Nov. 2005). “Foundations for engineering biology”. In: *Nature* 438(7067), pp. 449–453. ISSN: 1476-4687. DOI: 10.1038/nature04342.

- Erhel, S. and E. Jamet (Sept. 2013). “Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness”. In: *Computers & Education* 67, pp. 156–167. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2013.02.019.
- Esmaeili, Afshin et al. (Dec. 2015). “PROKARYO: an illustrative and interactive computational model of the lactose operon in the bacterium *Escherichia coli*”. In: *BMC Bioinformatics* 16(1). ISSN: 1471-2105. DOI: 10.1186/s12859-015-0720-z.
- Ferguson, R. et al. (2017). *Innovating Pedagogy 2017*. OpenUniv, pp. 1–48.
- Feurzeig, Wallace et al. (1969). “Programming-Languages as a Conceptual Framework for Teaching Mathematics. Final Report on the First Fifteen Months of the LOGO Project.” In:
- Francillette, Yannick, Abdelkader Gouaich, and Lyliya Abrouk (Aug. 2017). “Adaptive gameplay for mobile gaming”. In: *IEEE*, pp. 80–87. ISBN: 978-1-5386-3233-8. DOI: 10.1109/CIG.2017.8080419.
- Freitas, Sara de and Paul Maharg (2011). *Digital Games and Learning*. Continuum Press: London, UK.
- Galarneau, Lisa L. (2005). “Authentic Learning Experiences Through Play: Games, Simulations and the Construction of Knowledge”. In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.810065.
- Gardner, Timothy S, Charles R Cantor, and James J Collins (2000). “Construction of a genetic toggle switch in *Escherichia coli*”. In: 403, p. 4.
- Gentry, DianaMicks (May 15, 2013). *Materials Manufactured from 3D Printed Synthetic Biology Arrays*.
- Gibson, D. G. et al. (Feb. 29, 2008). “Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome”. In: *Science* 319(5867), pp. 1215–1220. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1151721.
- Girard, C., J. Ecalle, and A. Magnan (June 1, 2013). “Serious games as new educational tools: how effective are they? A meta-analysis of recent studies”. In: *Journal of Computer Assisted Learning* 29(3), pp. 207–219. ISSN: 1365-2729. DOI: 10.1111/j.1365-2729.2012.00489.x.
- Granic, Isabela, Adam Lobel, and Rutger C. M. E. Engels (2014). “The benefits of playing video games”. In: *American Psychologist* 69(1), pp. 66–78. ISSN: 1935-990X(Electronic),0003-066X(Print). DOI: 10.1037/a0034857.
- Greller, Wolfgang and Hendrik Drachslar (2012). “Translating Learning into Numbers: A Generic Framework for Learning Analytics”. In: *Journal of Educational Technology & Society* 15(3), pp. 42–57. ISSN: 1176-3647.
- Guardiola, Emmanuel (2016). “The Gameplay Loop: a Player Activity Model for Game Design and Analysis”. In: *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology - ACE2016*. the 13th International Conference. ACM Press: Osaka, Japan, pp. 1–7. ISBN: 978-1-4503-4773-0. DOI: 10.1145/3001773.3001791.
- H. Riedel-Kruse, Ingmar et al. (2011). “Design, engineering and utility of biotic games”. In: *Lab on a Chip* 11(1), pp. 14–22. DOI: 10.1039/C0LC00399A.

- Harvey, Hayden et al. (Nov. 1, 2014). “Innocent Fun or “Microslavery”?” In: *Hastings Center Report* 44(6), pp. 38–46. ISSN: 1552-146X. DOI: 10.1002/hast.386.
- Henneman, Lidewij et al. (Aug. 2013). “Public attitudes towards genetic testing revisited: comparing opinions between 2002 and 2010”. In: *European Journal of Human Genetics* 21(8), pp. 793–799. ISSN: 1476-5438. DOI: 10.1038/ejhg.2012.271.
- Himmelstein, Jesse, Mikael Couzic, et al. (2014). “RedWire: a novel way to create and re-mix games”. In: *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. the first ACM SIGCHI annual symposium. ACM Press: Toronto, Ontario, Canada, pp. 423–424. ISBN: 978-1-4503-3014-5. DOI: 10.1145/2658537.2661315.
- Himmelstein, Jesse, Raphael Goujet, et al. (2016). “Improving Citizen Science Games through Open Analytics Data”. In: *Human Computation* 3(1), pp. 119–141.
- Hodent, Celia (2017). *The Gamer’s Brain: How Neuroscience and UX Can Impact Video Game Design*. CRC Press.
- Huizinga, Johan (1938). “Homo ludens”. In: *Paris, Gallimard*.
- Hunicke, Robin (2005). “The Case for Dynamic Difficulty Adjustment in Games”. In: *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*. ACE '05. ACM: New York, NY, USA, pp. 429–433. ISBN: 978-1-59593-110-8. DOI: 10.1145/1178477.1178573.
- Hutchison, Clyde A., Ray-Yuan Chuang, et al. (Mar. 25, 2016). “Design and synthesis of a minimal bacterial genome”. In: *Science* 351(6280), aad6253. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aad6253.
- Hutchison, Clyde A., Scott N. Peterson, et al. (Dec. 10, 1999). “Global Transposon Mutagenesis and a Minimal Mycoplasma Genome”. In: *Science* 286(5447), pp. 2165–2169. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.286.5447.2165.
- Ingersoll, Richard M. (2003). *Is There Really a Teacher Shortage?: (382722004-001)*. type: dataset. American Psychological Association. DOI: 10.1037/e382722004-001.
- Jacobs, Ruud S, Jeroen Jansz, and Teresa de la Hera CondePumpido (2017). “The Key Features of Persuasive Games”. In: *New Perspectives on the Social Aspects of Digital Gaming: Multiplayer 2*.
- Jennett, Charlene, Ioanna Iacovides, et al. (2013). “Gamification in Citizen Cyber-science: Projects in Particle Physics and Synthetic Biology”. In: *Gamification 2013*. Waterloo, Canada.
- Jennett, Charlene, Laure Kloetzer, et al. (Dec. 31, 2016). “Creativity in Citizen Cyberscience”. In: *Human Computation* 3(1), pp. 181–204. ISSN: 2330-8001. DOI: 10.15346/hc.v3i1.10.
- Jenova Chen (2011). *Flow in Games, A Jenova Chen MFA Thesis*. Jenova Chen. URL: <http://www.jenovachen.com/flowingames/thesis.htm> (visited on 07/30/2018).

- Jewett, Michael C. and Anthony C. Forster (Oct. 2010). "Update on designing and building minimal cells". In: *Current opinion in biotechnology* 21(5), pp. 697–703. ISSN: 0958-1669. DOI: 10.1016/j.copbio.2010.06.008.
- Jr, C. Neal Stewart, Matthew D. Halfhill, and Suzanne I. Warwick (Oct. 2003). "Genetic modification: Transgene introgression from genetically modified crops to their wild relatives". In: *Nature Reviews Genetics* 4(10), pp. 806–817. ISSN: 1471-0064. DOI: 10.1038/nrg1179.
- Juul, Jesper (2003). "The Game, the Player, the World: Looking for a Heart of Gameness". In: *Level Up: Digital Games Research Conference Proceedings*. Level Up conference. Utrecht, p. 12.
- Juul, Jesper and Marleigh Norton (2009). "Easy to Use and Incredibly Difficult: On the Mythical Border Between Interface and Gameplay". In: *Proceedings of the 4th International Conference on Foundations of Digital Games*. FDG '09. ACM: New York, NY, USA, pp. 107–112. ISBN: 978-1-60558-437-9. DOI: 10.1145/1536513.1536539.
- Kang, Sean H. K. (Mar. 2016). "Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction". In: *Policy Insights from the Behavioral and Brain Sciences* 3(1). Ed. by Susan T. Fiske, pp. 12–19. ISSN: 2372-7322, 2372-7330. DOI: 10.1177/2372732215624708.
- Karr, Jonathan R. et al. (July 20, 2012). "A Whole-Cell Computational Model Predicts Phenotype from Genotype". In: *Cell* 150(2), pp. 389–401. ISSN: 0092-8674. DOI: 10.1016/j.cell.2012.05.044.
- Kawrykow, Alexander et al. (Mar. 7, 2012). "Phylo: A Citizen Science Approach for Improving Multiple Sequence Alignment". In: *PLOS ONE* 7(3), e31362. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0031362.
- Keese, Paul (July 2008). "Risks from GMOs due to Horizontal Gene Transfer". In: *Environmental Biosafety Research* 7(3), pp. 123–149. ISSN: 1635-7922, 1635-7930. DOI: 10.1051/ebr:2008014.
- Kiili, Kristian, Harri Ketamo, and Timo Lainema (2011). "Reflective Thinking in Games: Triggers and Constraints". In: *Leading Issues in Games Based Learning*, p. 178.
- Kirschner, Paul A. (Feb. 2002). "Cognitive load theory: implications of cognitive load theory on the design of learning". In: *Learning and Instruction* 12(1), pp. 1–10. ISSN: 09594752. DOI: 10.1016/S0959-4752(01)00014-7.
- Kitada, Tasuku et al. (2018). "Programming gene and engineered-cell therapies with synthetic biology". In: *Science* 359(6376), eaad1067.
- Kleij, Fabienne M. van der et al. (Jan. 2012). "Effects of feedback in a computer-based assessment for learning". In: *Computers & Education* 58(1), pp. 263–272. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2011.07.020.
- Klopfer, Eric and Scot Osterweil (2013). "The Boom and Bust and Boom of Educational Games". In: *Transactions on Edutainment IX*. Ed. by Zhigeng Pan et al. Red. by David Hutchison et al. Vol. 7544. Springer Berlin Heidelberg: Berlin, Heidelberg, pp. 290–296. ISBN: 978-3-642-37041-0 978-3-642-37042-7. DOI: 10.1007/978-3-642-37042-7\_21.

- Knight, Thomas (2003). “Idempotent vector design for standard assembly of bio-bricks”. In:
- Knight, Thomas F. (2005). *Engineering novel life*. undefined. URL: /paper/Engineering-novel-life-Knight/7f245ae11fb7a66cd833891ddf01472f7ef6a1cb (visited on 09/03/2018).
- Kolb, Alice Y. and David A. Kolb (Feb. 16, 2010). “Learning to play, playing to learn: A case study of a ludic learning space”. In: *Journal of Organizational Change Management* 23(1). Ed. by Charalampos Mainemelis, pp. 26–50. ISSN: 0953-4814. DOI: 10.1108/09534811011017199.
- Laurel, Brenda (2002). “Utopian Entrepreneur”. In: *Utopian Studies* 13(1), pp. 214–216.
- Ledford, Heidi (Oct. 7, 2010). “Garage biotech: Life hackers”. In: *Nature* 467(7316), pp. 650–652. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/467650a.
- Lee, Jeehyung et al. (Jan. 24, 2014). “RNA design rules from a massive open laboratory”. In: *Proceedings of the National Academy of Sciences*, p. 201313039. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1313039111.
- Lee, Sung Kuk et al. (Dec. 1, 2008). “Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels”. In: *Current Opinion in Biotechnology*. Chemical biotechnology / Pharmaceutical biotechnology 19(6), pp. 556–563. ISSN: 0958-1669. DOI: 10.1016/j.copbio.2008.10.014.
- Levinthal, Cyrus (1969). “How to fold graciously”. In: *Mossbauer spectroscopy in biological systems* 67, pp. 22–24.
- Lin, Yanni et al. (June 17, 2014). “CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences”. In: *Nucleic Acids Research* 42(11), pp. 7473–7485. ISSN: 0305-1048. DOI: 10.1093/nar/gku402.
- Liu, Tiangang and Chaitan Khosla (2010). “Genetic engineering of Escherichia coli for biofuel production”. In: *Annual review of genetics* 44, pp. 53–69.
- Macklin, Derek N, Nicholas A Ruggero, and Markus W Covert (Aug. 2014). “The future of whole-cell modeling”. In: *Current Opinion in Biotechnology* 28, pp. 111–115. ISSN: 09581669. DOI: 10.1016/j.copbio.2014.01.012.
- Mader, Stéphanie, Stéphane Natkin, and Guillaume Levieux (2012). “How to Analyse Therapeutic Games: The Player / Game / Therapy Model”. In: *Entertainment Computing - ICEC 2012*. Ed. by Marc Herrlich, Rainer Malaka, and Maic Masuch. Vol. 7522. Springer Berlin Heidelberg: Berlin, Heidelberg, pp. 193–206. ISBN: 978-3-642-33541-9 978-3-642-33542-6. DOI: 10.1007/978-3-642-33542-6\_17.
- Malyshev, Denis A. et al. (May 2014). “A semi-synthetic organism with an expanded genetic alphabet”. In: *Nature* 509(7500), pp. 385–388. ISSN: 1476-4687. DOI: 10.1038/nature13314.
- Marchand, André and Thorsten Hennig-Thurau (Aug. 2013). “Value Creation in the Video Game Industry: Industry Economics, Consumer Benefits, and Research Opportunities”. In: *Journal of Interactive Marketing* 27(3), pp. 141–157. ISSN: 10949968. DOI: 10.1016/j.intmar.2013.05.001.

- Marris, Claire (Jan. 2, 2015). “The Construction of Imaginaries of the Public as a Threat to Synthetic Biology”. In: *Science as Culture* 24(1), pp. 83–98. ISSN: 0950-5431, 1470-1189. DOI: 10.1080/09505431.2014.986320.
- Maudet, C. et al. (Mar. 1, 2002). “Microsatellite DNA and recent statistical methods in wildlife conservation management: applications in Alpine ibex [Capra ibex (ibex)]”. In: *Molecular Ecology* 11(3), pp. 421–436. ISSN: 1365-294X. DOI: 10.1046/j.0962-1083.2001.01451.x.
- Mayes, J Terence (n.d.). “Learning technology and groundhog day”. In: ().
- Morisset, Thomas (2014). “Nature de la lecture et matérialité des livres dans les jeux vidéo”. In: *Mémoires du livre / Studies in Book Culture* 5(2). ISSN: 1920-602X. DOI: 10.7202/1024776ar.
- Murphy, Robert F. (2018). *What is Computational Biology? | Computational Biology Department*. Computational Biology Department | Carnegie Mellon University. URL: <http://www.cbd.cmu.edu/about-us/what-is-computational-biology/> (visited on 09/15/2018).
- Owens, Trevor (Dec. 1, 2012). “Teaching intelligent design or sparking interest in science? What players do with Will Wright’s Spore”. In: *Cultural Studies of Science Education* 7(4), pp. 857–868. ISSN: 1871-1502, 1871-1510. DOI: 10.1007/s11422-012-9383-5.
- Paddon, Chris J. and Jay D. Keasling (May 2014). “Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development”. In: *Nature Reviews Microbiology* 12(5), pp. 355–367. ISSN: 1740-1534. DOI: 10.1038/nrmicro3240.
- Paddon, Christopher J et al. (2013). “High-level semi-synthetic production of the potent antimalarial artemisinin”. In: *Nature* 496(7446), p. 528.
- Papastergiou, Marina (Jan. 2009). “Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation”. In: *Computers & Education* 52(1), pp. 1–12. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2008.06.004.
- Papert, Seymour (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, Inc.: New York, NY, USA. ISBN: 978-0-465-04627-0.
- Paras, Brad and Jim Bizzocchi (2005). “Game, Motivation, and Effective Learning: An Integrated Model for Educational Game Design”. In: p. 8.
- Pashler, Harold et al. (Dec. 2008). “Learning Styles: Concepts and Evidence”. In: *Psychological Science in the Public Interest* 9(3), pp. 105–119. ISSN: 1529-1006, 1539-6053. DOI: 10.1111/j.1539-6053.2009.01038.x.
- Paul, Christopher A. (May 2011). “Optimizing Play: How Theocracy Changes Gameplay and Design”. In: *Game Studies* 11(2). ISSN: 1604-7982.
- Pauwels, Eleonore (Feb. 2013). “Public Understanding of Synthetic Biology”. In: *BioScience* 63(2), pp. 79–89. ISSN: 1525-3244, 0006-3568. DOI: 10.1525/bio.2013.63.2.4.
- Pearce, F. (2003). “Going bananas”. In: *New Scientist*, pp. 26–29.

- Perron, Bernard and Dominic Arsenaault (2008). “In the frame of the magic cycle: The circle (s) of gameplay”. In: *The video game theory reader 2*. Routledge, pp. 131–154.
- Piaget, Jean (1970). *Psychologie et épistémologie*. Vol. 73. Paris, Gonthier.
- Prensky, Marc (2001). “Digital natives, digital immigrants part 1”. In: *On the horizon* 9(5), pp. 1–6.
- Prensky, Marc (2003). “Digital game-based learning”. In: *Computers in Entertainment (CIE)* 1(1), pp. 21–21.
- Purnick, Priscilla E. M. and Ron Weiss (June 2009). “The second wave of synthetic biology: from modules to systems”. In: *Nature Reviews Molecular Cell Biology* 10(6), pp. 410–422. ISSN: 14710072. DOI: 10.1038/nrm2698.
- Radzicka, A. and R. Wolfenden (Jan. 6, 1995). “A proficient enzyme”. In: *Science* 267(5194), pp. 90–93. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.7809611.
- Ramanauskaitė, Eglė Marija and Mordechai Haklay (Dec. 31, 2016). “Creativity and Learning in Citizen Cyberscience – Lessons from the Citizen Cyberlab Summit”. In: *Human Computation* 3(1), pp. 5–24. ISSN: 2330-8001. DOI: 10.15346/hc.v3i1.3.
- Reynolds, R., R. M. Bermúdez-Cruz, and M. J. Chamberlin (Mar. 1992). “Parameters affecting transcription termination by Escherichia coli RNA polymerase. I. Analysis of 13 rho-independent terminators.” In: *Journal of molecular biology* 224(1), pp. 31–51. ISSN: 0022-2836. DOI: 10.1016/0022-2836(92)90574-4.
- Rollings, Andrew and Ernest Adams (2003). *Andrew Rollings and Ernest Adams on game design*. New Riders.
- Ryan, Marie-Laure, Lori Emerson, and Benjamin J Robertson (2014). *The Johns Hopkins guide to digital media*. JHU Press.
- Sanchez, Eric (Nov. 2014). “Le paradoxe du marionnettiste”. Habilitation à diriger des recherches. Université Paris 5 Sorbonne Descartes.
- Sanchez, Éric, Valérie Martinez-Emin, and Nadine Mandran (2015). “Jeu-game, jeu-play, vers une modélisation du jeu. Une étude empirique à partir des traces numériques d’interaction du jeu Tamagocours”. In: *Sciences et Technologies de l’Information et de la Communication pour l’Éducation et la Formation* 22(1), pp. 9–44. DOI: 10.3406/stice.2015.1685.
- Sanchez, Éric, Muriel Ney, and Jean-Marc Labat (2011). “Jeux sérieux et pédagogie universitaire : de la conception à l’évaluation des apprentissages”. In: *Revue internationale des technologies en pédagogie universitaire* 8(1), p. 48. ISSN: 1708-7570. DOI: 10.7202/1005783ar.
- Schmidt, Markus, Agomoni Ganguli-Mitra, et al. (Oct. 10, 2009). “A priority paper for the societal and ethical aspects of synthetic biology”. In: *Systems and Synthetic Biology* 3(1), p. 3. ISSN: 1872-5333. DOI: 10.1007/s11693-009-9034-7.
- Schmidt, Markus, Olga Radchuk, and Camillo Meinhart (2014). “A Serious Game for Public Engagement in Synthetic Biology”. In: *Games for Training, Education, Health and Sports*. Ed. by Stefan Göbel and Josef Wiemeyer. Red. by David Hutchison et al. Vol. 8395. Springer International Publishing: Cham,

- pp. 77–85. ISBN: 978-3-319-05971-6 978-3-319-05972-3. DOI: 10.1007/978-3-319-05972-3\_9.
- Schneider, Maria Victoria and Rafael C. Jimenez (Dec. 27, 2012). “Teaching the Fundamentals of Biological Data Integration Using Classroom Games”. In: *PLOS Computational Biology* 8(12), e1002789. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002789.
- Schrader, Peter, Hasan Deniz, and Joshua Keilty (2016). “Breaking SPORE: Building Instructional Value in Science Education using a Commercial, Off-the Shelf Game”. In: p. 11.
- Séralini, Gilles-Eric et al. (Jan. 2014). “Retraction notice to “Long term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize” [Food Chem. Toxicol. 50 (2012) 4221–4231]”. In: *Food and Chemical Toxicology* 63, p. 244. ISSN: 02786915. DOI: 10.1016/j.fct.2013.11.047.
- Serrano-Laguna, Ángel et al. (2012). “Tracing a Little for Big Improvements: Application of Learning Analytics and Videogames for Student Assessment”. In: *Procedia Computer Science*. 4th International Conference on Games and Virtual Worlds for Serious Applications(VS-GAMES’12) 15, pp. 203–209. ISSN: 1877-0509. DOI: 10.1016/j.procs.2012.10.072.
- Shachrai, Irit et al. (June 11, 2010). “Cost of Unneeded Proteins in E. coli Is Reduced after Several Generations in Exponential Growth”. In: *Molecular Cell* 38(5), pp. 758–767. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2010.04.015.
- Sharp, Laura A. (2012). “Stealth Learning: Unexpected Learning Opportunities through Games”. In: *Journal of Instructional Research* 1, pp. 42–48. ISSN: 2159-0281.
- Shute, Valerie J (2011). “Stealth assessment in computer-based games to support learning”. In: *Computer games and instruction* 55(2), pp. 503–524.
- Shute, Valerie J., Eric G. Hansen, and Russell G. Almond (Dec. 2008). “You Can’T Fatten A Hog by Weighing It - Or Can You? Evaluating an Assessment for Learning System Called ACED”. In: *Int. J. Artif. Intell. Ed.* 18(4), pp. 289–316. ISSN: 1560-4292.
- Sicart, Miguel (2008). “Newsgames: Theory and Design”. In: *Entertainment Computing - ICEC 2008*. Ed. by Scott M. Stevens and Shirley J. Saldamarco. Vol. 5309. Springer Berlin Heidelberg: Berlin, Heidelberg, pp. 27–33. ISBN: 978-3-540-89221-2 978-3-540-89222-9. DOI: 10.1007/978-3-540-89222-9\_4.
- Siegele, D. A. and J. C. Hu (July 22, 1997). “Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations”. In: *Proceedings of the National Academy of Sciences* 94(15), pp. 8168–8172. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.94.15.8168.
- Singer, Dorothy G. et al. (Aug. 24, 2006). *Play = Learning: How Play Motivates and Enhances Children’s Cognitive and Social-Emotional Growth*. Google-Books-ID: 9NFIuO7WrdwC. Oxford University Press, USA. 290 pp. ISBN: 978-0-19-530438-1.
- Stager, Gary S (2016). *Seymour Papert (1928–2016)*.

- Stegman, Melanie (2014). “Immune Attack players perform better on a test of cellular immunology and self confidence than their classmates who play a control video game”. In: *Faraday Discussions* 169(0), pp. 403–423. DOI: 10.1039/C4FD00014E.
- Steinkuehler, Constance and Sean Duncan (Dec. 2008). “Scientific Habits of Mind in Virtual Worlds”. In: *Journal of Science Education and Technology* 17(6), pp. 530–543. ISSN: 1059-0145, 1573-1839. DOI: 10.1007/s10956-008-9120-8.
- Susi, Tarja, Mikael Johannesson, and Per Backlund (2007). *Serious Games : An Overview*. Institutionen för kommunikation och information.
- Sutcher, Leib, Linda Darling-Hammond, and Desiree Carver-Thomas (2016). “A coming crisis in teaching”. In: *Teacher supply, demand, and shortages in the US*.
- Sweller, J (1999). “Instructional Design in Technical Areas. Australian Education Review, No. 43. PCS Data Processing”. In: *Inc. USA*.
- Symes, Colin (July 2004). “A sound education: the gramophone and the classroom in the United Kingdom and the United States, 1920-1940”. In: *British Journal of Music Education* 21(2), pp. 163–178. ISSN: 0265-0517, 1469-2104. DOI: 10.1017/S0265051704005674.
- Systems biology - Latest research and news / Nature* (2018). URL: <https://www.nature.com/subjects/systems-biology> (visited on 09/15/2018).
- Tang, Stephen and Dr Martin Hanneghan (2007). “Describing Games for Learning: Terms, Scope and Learning Approaches”. In: p. 5.
- Teusink, Bas et al. (Sept. 1, 2000). “Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry”. In: *European Journal of Biochemistry* 267(17), pp. 5313–5329. ISSN: 1432-1033. DOI: 10.1046/j.1432-1327.2000.01527.x.
- Thomas, Pradeepa et al. (June 2012). “How to Evaluate Competencies in Game-Based Learning Systems Automatically?” In: *11th international conference on Intelligent Tutoring Systems*. Vol. 7315. Lecture Notes in Computer Science. Springer: Chania, Greece, pp. 168–173. DOI: 10.1007/978-3-642-30950-2\_22.
- Tricot, André et al. (Apr. 17, 2003). “Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH”. In: *Environnements Informatiques pour l'Apprentissage Humain 2003*. ATIEF, pp. 391–402.
- Van Eck, Richard (2006). “Digital game-based learning: It's not just the digital natives who are restless”. In:
- Villa-Komaroff, L. et al. (Aug. 1, 1978). “A bacterial clone synthesizing proinsulin”. In: *Proceedings of the National Academy of Sciences* 75(8), pp. 3727–3731. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.75.8.3727.
- Wang, Feng and Michael J. Hannafin (2005). “Design-Based Research and Technology-Enhanced Learning Environments”. In: *Educational Technology Research and Development* 53(4), pp. 5–23.

- Washington, Peter et al. (Sept. 7, 2018). “An interactive programming paradigm for realtime experimentation with remote living matter”. In: *bioRxiv*, p. 236919. DOI: 10.1101/236919.
- Weiß, Andrea Y. et al. (Mar. 3, 2015). “Mechanistic links between cellular trade-offs, gene expression, and growth”. In: *Proceedings of the National Academy of Sciences* 112(9), E1038–E1047. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1416533112.
- Westera, Wim, Rob Nadolski, and Hans Hummel (Sept. 2, 2014). “Serious Gaming Analytics: What Students’ Log Files Tell Us about Gaming and Learning”. In: Wilkinson, Nathan, Rebecca P. Ang, and Dion H. Goh (July 1, 2008). “Online Video Game Therapy for Mental Health Concerns: A Review”. In: *International Journal of Social Psychiatry* 54(4), pp. 370–382. ISSN: 0020-7640. DOI: 10.1177/0020764008091659.
- Willingham, Daniel T., Elizabeth M. Hughes, and David G. Dobolyi (July 2015). “The Scientific Status of Learning Styles Theories”. In: *Teaching of Psychology* 42(3), pp. 266–271. ISSN: 0098-6283, 1532-8023. DOI: 10.1177/0098628315589505.
- Wolf, Mark JP (2001). *The medium of the video game*. University of Texas Press.
- Wortel, Meike T. et al. (July 6, 2016). “Evolutionary pressures on microbial metabolic strategies in the chemostat”. In: *Scientific Reports* 6, p. 29503. ISSN: 2045-2322. DOI: 10.1038/srep29503.
- Ye, Xudong et al. (Jan. 14, 2000). “Engineering the Provitamin A ( $\beta$ -Carotene) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm”. In: *Science* 287(5451), pp. 303–305. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.287.5451.303.



# Appendix A

## Annex 1: tables

#	Title	Code
01	[Timestamp]	[Timestamp]
02	Are you interested in learning more about...	Want to learn more about Biology
03	Are you interested in learning more about...	Want to learn more about Synthetic biology
04	Are you interested in learning more about...	Want to learn more about Video games
05	Are you interested in learning more about...	Want to learn more about Engineering
06	Have you ever played Hero.Coli?	Played Hero.Coli
07	How old are you?	Age
08	What is your gender?	Gender
09	Are you interested in video games?	Interested in video games
10	Are you interested in biology?	Interested in biology
11	How long have you studied biology?	Studied biology
12	Do you play video games?	Play video games
13	Have you ever heard about synthetic biology or BioBricks, outside of Hero.Coli?	Heard about Synthetic biology or BioBricks
14	Do you volunteer to contribute to our study by answering 9 more questions? (5 min)	Volunteered to answer more questions
15	Did you enjoy playing the game?	Enjoyed playing
16	In order to modify the abilities of the bacterium, you have to...	Genotype and phenotype
17	What are BioBricks and devices?	BioBricks and devices composition
18	Find the antibiotic:	Ampicillin antibiotic
19	Plasmid is...	Name: Plasmid
20	Represents the end of a device...	Function: TER
21	Promoter is...	Name: PR

22	Represents the ability given...	Function - game: CDS
23	Terminator is...	Name: TER
24	Codes a protein...	Function - biology: CDS
25	RBS is...	Name: RBS
26	Can represent GFP...	Example: CDS
27	Coding Sequence is...	Name: CDS
28	Controls when the device is active...	Function: PR
29	Controls the level of expression, and thus how much the ability will be affected...	Function: RBS
30	Makes it possible to equip an additional device.	Function: Plasmid
31	Operator is...	Name: Operator XXX
32	What does this device do?	Device: RBS:PCONS:FLHDC:TER XXX
33	What does this device do?	Device: PCONS:RBS:FLHDC:TER
34	What does this device do?	Device: PBAD:RBS:GFP:TER
35	What does this device do?	Device: PBAD:GFP:RBS:TER XXX
36	What does this device do?	Device: GFP:RBS:PCONS:TER XXX
37	What does this device do?	Device: PCONS:GFP:RBS:TER XXX
38	What does this device do?	Device: AMPR:RBS:PCONS:TER XXX
39	What does this device do?	Device: RBS:PCONS:AMPR:TER XXX
40	When does green fluorescence happen?	Green fluorescence
41	What happens when you unequip the movement device?	Unequip the movement device: effect
42	Last question. Next page only contains remarks. Guess: you have crafted a functional device containing an arabinose-induced promoter and an arabinose Coding Sequence (CDS). What will happen?	Device: PBAD:RBS:ARA:TER
43	You can write down remarks here.	Remarks
44	[userId]	[userId]
45	[Language]	[Language]

46	[Temporality]	[Temporality]
----	---------------	---------------

Table A.3: Questions: number, title, code

Date	Milestone
2012	Citizen Cyberlab funding starts
2012-09 to 2013-03	first design by a CRI team comprising the Digital SB club, synthetic biologists, a game designer, and a 2d artist
2013-03 to 2013-07	implementation of the biochemical simulator
2013-06 to 2013-08	prototype development using Unity
2013-08 to 2013-12	development for iGAM4ER 2013; basic craft
2013-12 to 2014-10	better interface, localized text - French/English; genetic crafting; parametrization through xml files
2014-10	game version 1.12: playtest at Cité des Sciences - see section 4.2.4
2014	Citizen Cyberlab funding ends - IDEFI IIFR funding starts
<b>2014-11</b>	<b>PhD starts</b>
2014-11 to 2015-05	bugfixing following the playtest at the Cité des Sciences; refactoring, cleaning; game menu; sandbox mode; biochemical simulator with new bricks and functional genetic crafting; RedMetrics integration, respawn as cell-division
2015-04	<i>first RedMetrics event logged</i>
2015-05 to 2016-05	runtime performance optimizations; code cleaning; update of project and of subcomponents; further RedMetrics integration
2015-09	link to MOOCs added
2015-12	link to Google form
2016-05 to 2016-09	game version 1.50: new map; major gameplay changes; re-designed tutorial
2016-09	<i>first HTML5 build</i>
2016-09 to 2017-05	polishing, bugfixing of new tutorials
2017-05-13	first answers to the updated online survey
2017-05 to 2017-06	arcade support, including sounds and new music
2017-06	main development on the game are halted; from then on, only bugfixes or minor changes needed for external systems - website, survey, analytics - have been implemented.
2017-09 to 2018-04	development of analytics tools in Python
2018-04-10 to 2018-04-28	game version 1.60: final experiment in Cité des Sciences museum, Paris; publication of current Hero.Coli version

Table A.1: Timeline of development of Hero.Coli

CPU	Intel Core i7-4770
GPU	NVidia GeForce GTX 760
Memory	16Gb DDR3 SDRAM
Hard Drive	1Tb Seagate Constellation ES.3
OS	Linux 64 bits on 4 gaming stations, 32 on the 5th. On all machines there was a dual boot with Windows 7 but the standalone version we built especially for this test worked better on Linux.
Headphones	CM Storm Sonuz
Mouse	CM Storm Inferno
Keyboard	Logitech Gaming Keyboard G510
Screen	LG 27EA83-D 27"

Table A.2: April 2018 Cite des Sciences Experiment: PC configuration



# Appendix B

## Annex 2: graphs

## B.1 Figures referenced in section 6.2.2

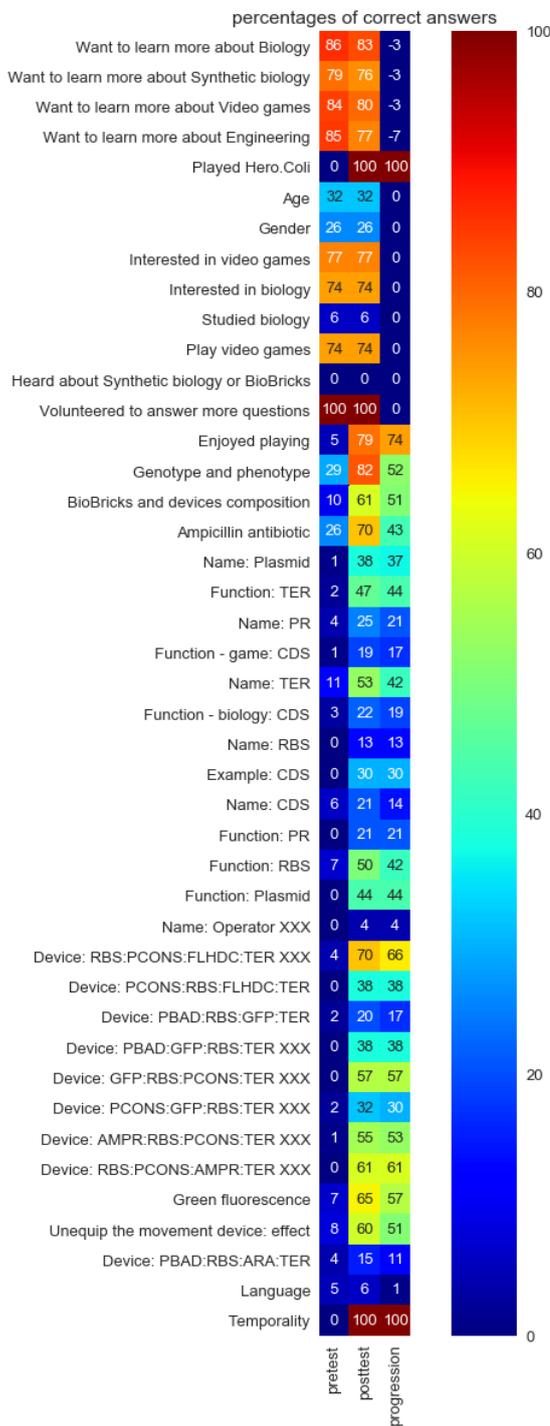


Figure B.1:

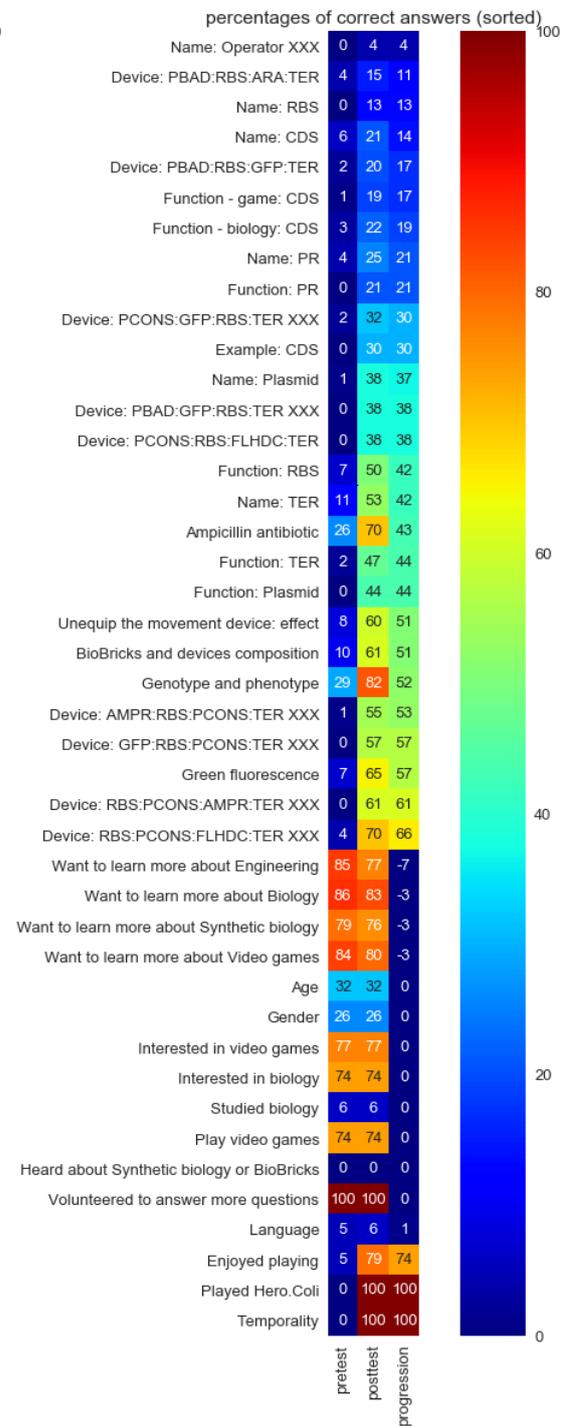


Figure B.2:

Figure B.3: B.1 Percentages of positive answers in pretest, posttest, and percentage increase, B.2 sorted by increase (enlarged)

## **B.2 Figures referenced in sections 6.1.3 and 6.1.4**

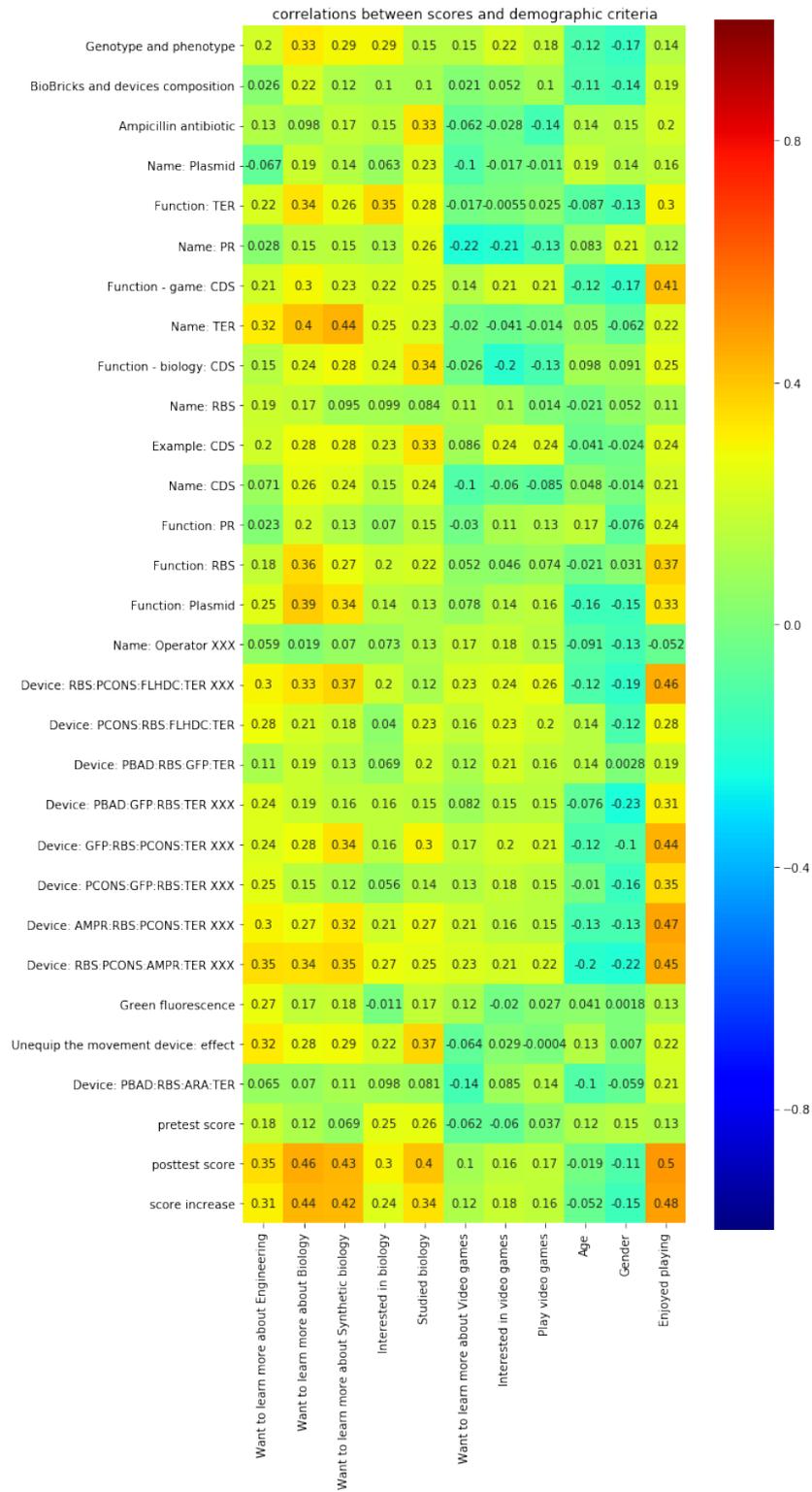


Figure B.4: Matrix of correlations of demographic features and interests against scores (enlarged)

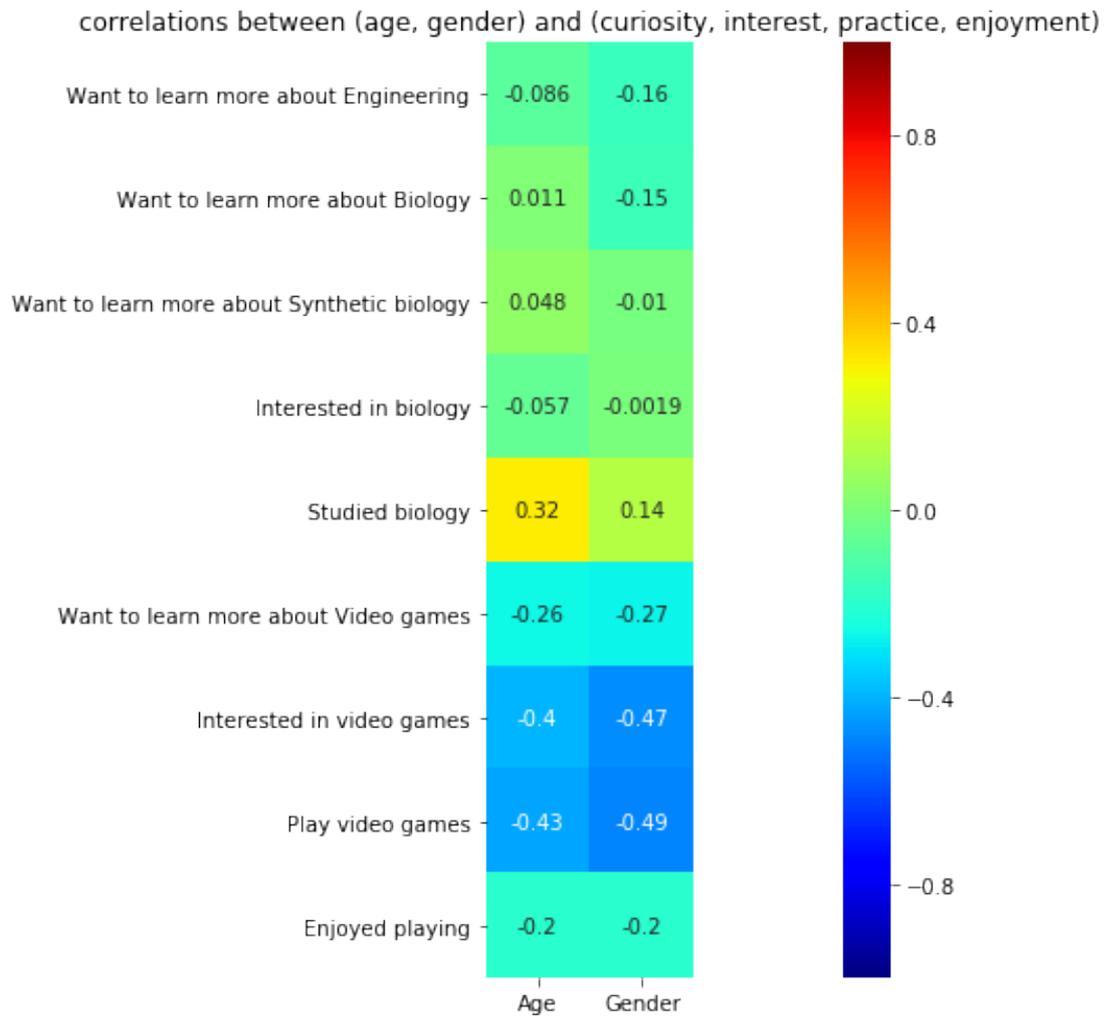


Figure B.5: Matrix of correlations of participant demographic features against their curiosity, interests, and practice (enlarged)

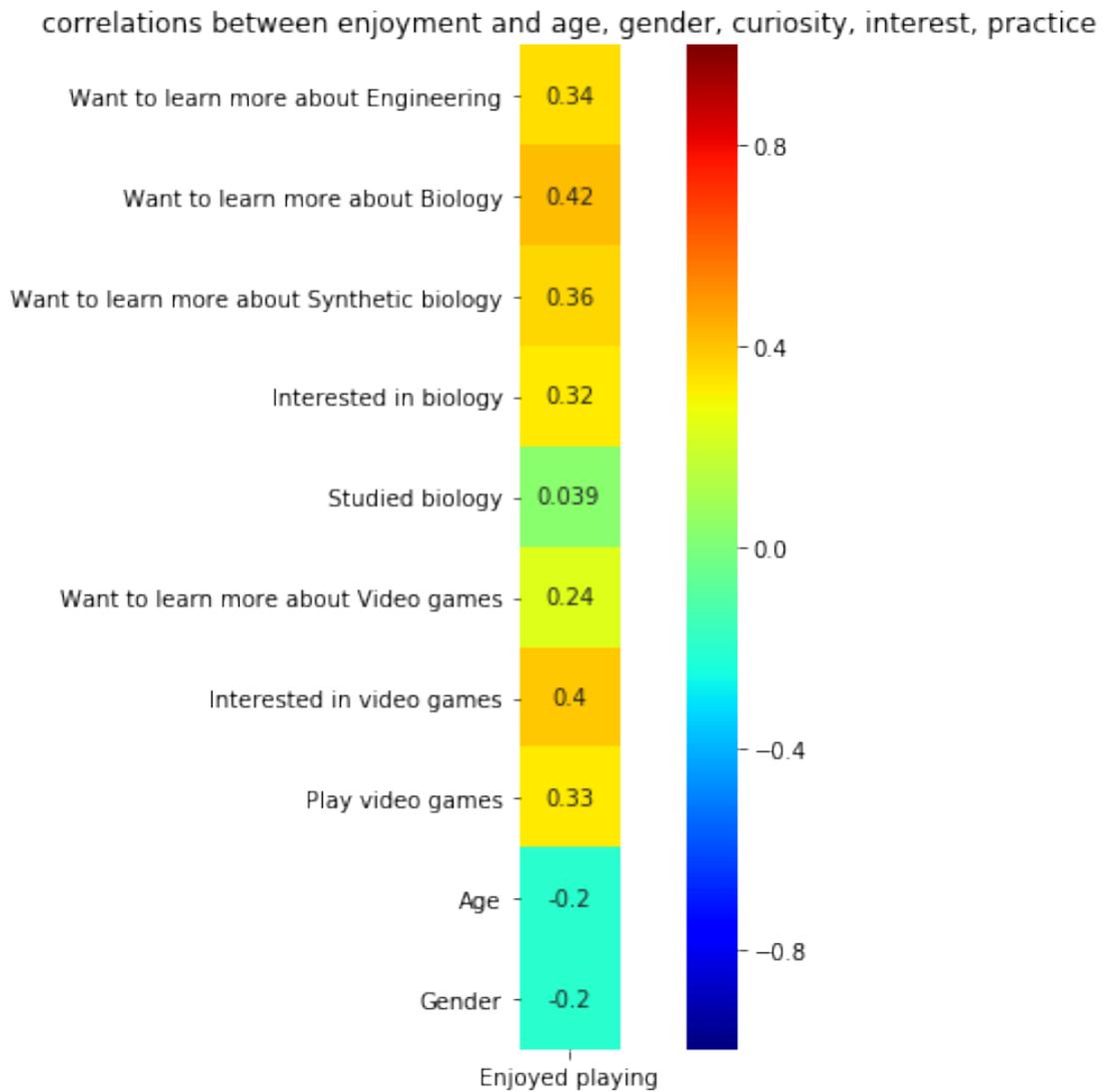


Figure B.6: Matrix of correlations of enjoyment against participants' characteristics (enlarged)

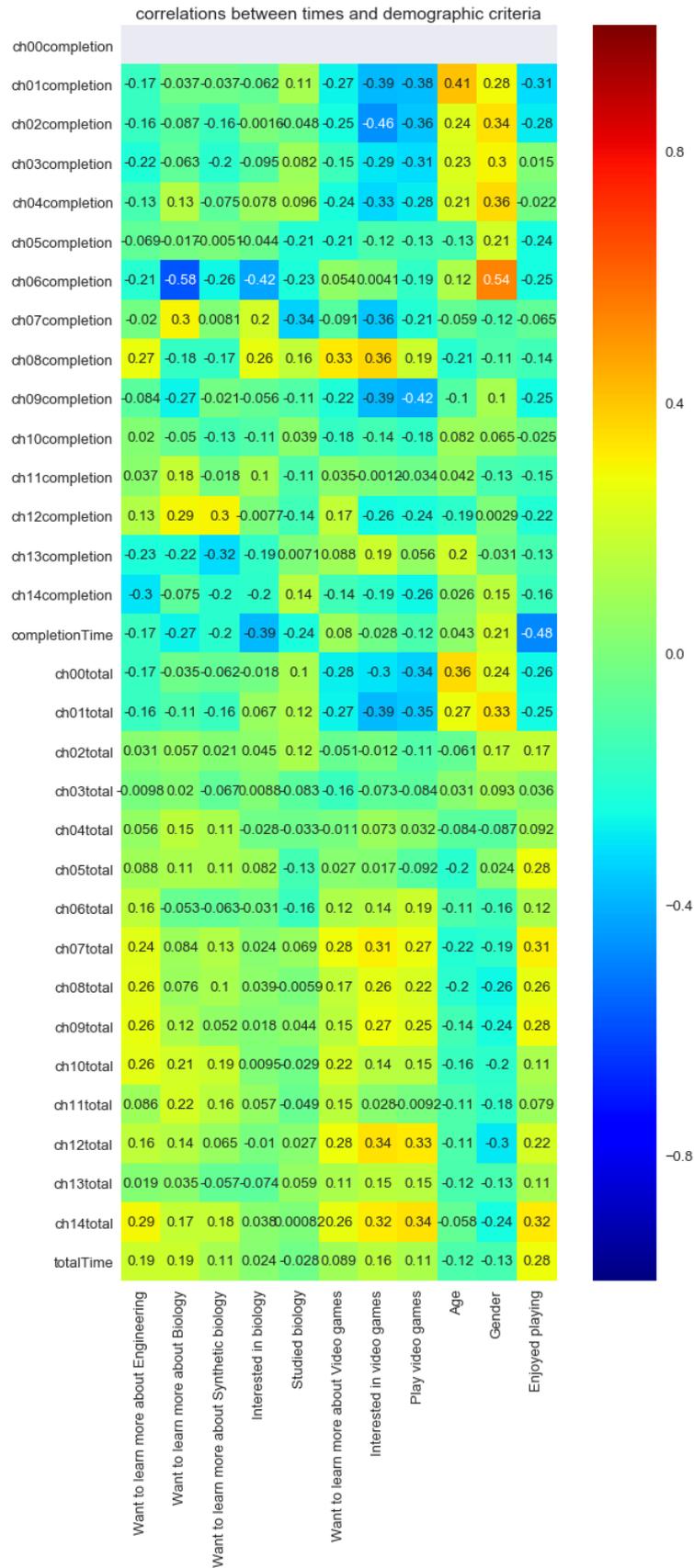


Figure B.7: Matrix of correlations of play times against participants' self-assessed data (enlarged)

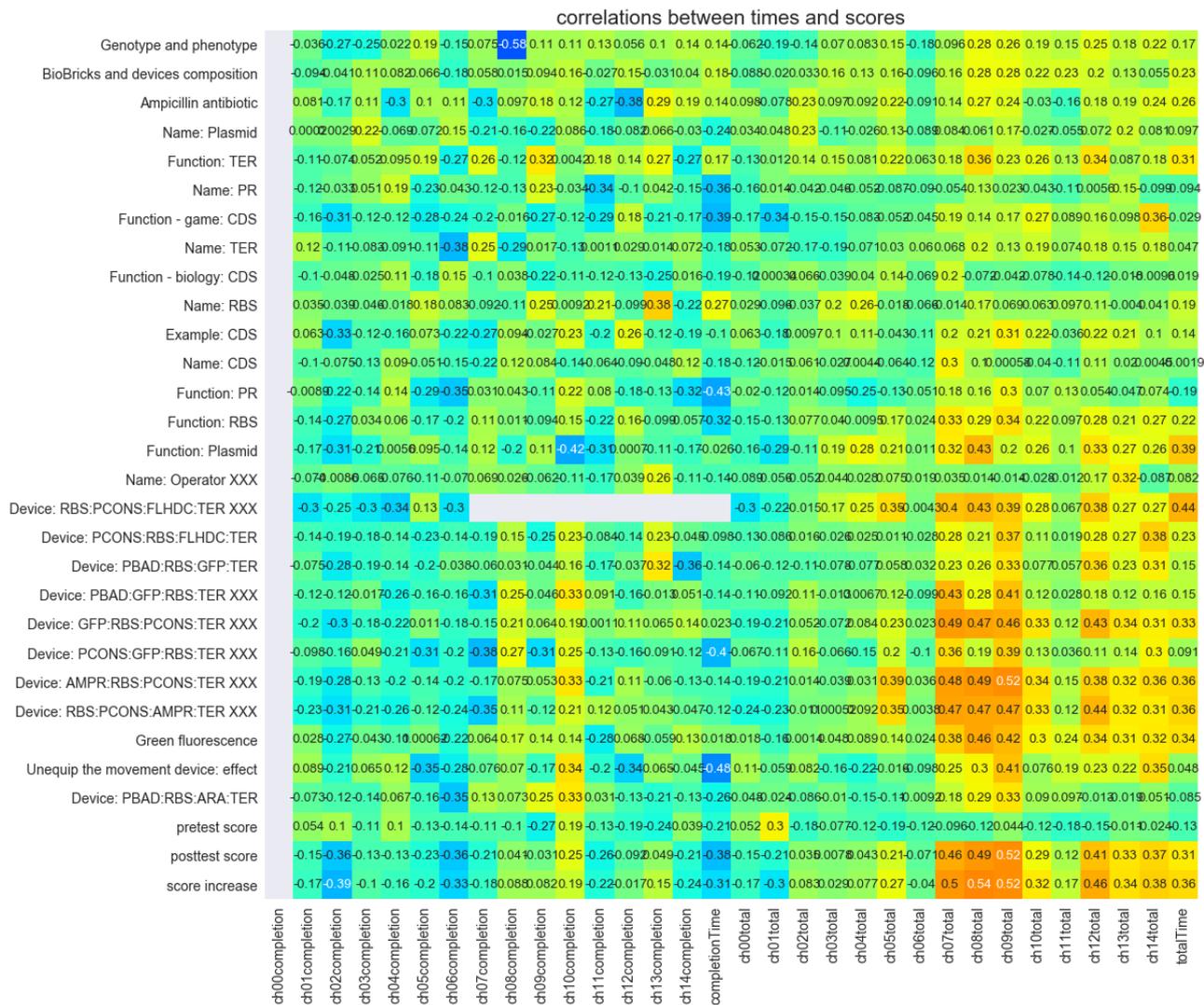


Figure B.8: Correlation matrix of the play times against score per question and total scores (enlarged)

# Appendix C

## Annex 3: surveys

# C.1 1.12

## Hero.Coli study (v0.2.en)

Put your answers for each question in the box below the question.

### What's your background?

---

1. How old are you?

---

2. What's your gender?

*Mark only one oval.*

male

female

other

3. What's your knowledge in biology?

*Mark only one oval.*

	1	2	3	4	5	6	7	
none	<input type="radio"/>	expert						

4. What's your level of interest in biology?

*Mark only one oval.*

	1	2	3	4	5	
not interested	<input type="radio"/>	enthusiast				

5. What's your level of interest in games?

*Mark only one oval.*

	1	2	3	4	5	
not interested	<input type="radio"/>	enthusiast				

6. How many times have you played this video game, Hero.Coli?

*Mark only one oval.*

never

first time

twice to five times

more than five times

### What have you understood?

---

7. How scientifically accurate did you find the game?

*Mark only one oval.*

not accurate

accurate

I don't know

**8. In the game, which one is a genetic device?**

(differentiate BioBrick and device)  
*Mark only one oval.*

- Hyperflagellation
- RBS 23%
- RFP ORF

**9. Spot the non-BioBrick:**

(differentiate BioBrick and device)  
*Mark only one oval.*

- GFP
- Terminator
- RBS

**10. Spot the correct sequence:**

*Mark only one oval.*

- Promoter-ORF-RBS-Terminator
- RBS-Promoter-ORF-Terminator
- Promoter-RBS-ORF-Terminator

**11. Which biobrick controls only the efficiency - level of expression - of the genetic device?**

*Mark only one oval.*

- ORF
- RBS
- Promoter

**12. Which biobrick controls what is produced by the genetic device?**

*Mark only one oval.*

- RBS
- Promoter
- ORF

**13. When does GFP emit green light?**

*Mark only one oval.*

- Near green rocks
- Always
- Never
- When receiving blue light

## **What have you understood in biology in general?**

Cellia (the bacterium in the game) is an E. coli. How realistic is the game compared to real-life E. coli?

The following questions refer to real-life E. coli:

**14. E. coli can move through the use of flagella.**

*Mark only one oval.*

- no
- yes
- I don't know

15. **E. coli can move by crawling like a snake.**

*Mark only one oval.*

- no
- yes
- I don't know

16. **E. coli has eyes.**

*Mark only one oval.*

- no
- yes
- I don't know

17. **E. coli looks like a tadpole.**

*Mark only one oval.*

- no
- yes
- I don't know

18. **E. coli divides.**

*Mark only one oval.*

- no
- yes
- I don't know

19. **E. coli wanders around to explore and harvest resources.**

*Mark only one oval.*

- no
- yes
- I don't know

20. **E. coli mainly evolves by absorbing foreign dna.**

*Mark only one oval.*

- no
- yes
- I don't know

21. **An isolated E. coli evolves by getting stronger little by little.**

*Mark only one oval.*

- no
- yes
- I don't know

22. **E. coli can be killed by some chemicals.**

*Mark only one oval.*

- no
- yes
- I don't know

23. **E. coli lives in aquatic environments.**

*Mark only one oval.*

- no
- yes
- I don't know

24. **E. coli struggle against bubbles and rocks.**

*Mark only one oval.*

- no
- yes
- I don't know

25. **E. coli can glow green by producing GFP when exposed to blue light.**

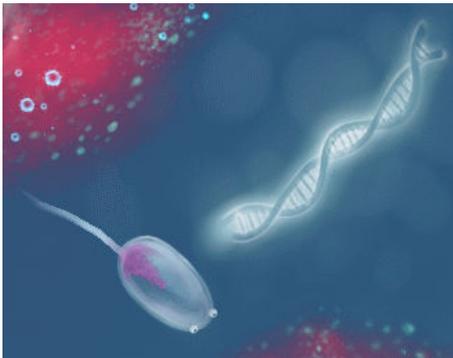
*Mark only one oval.*

- no
- yes
- I don't know

26. **The relative sizes of DNA and E. coli are accurate in the picture below.**

*Mark only one oval.*

- no
- yes
- I don't know



# C.2 1.50 - 2016-06 to 2017-06

## Hero.Coli survey

Your answers will help us improve the game! Thank you for your participation.

### What's your background?

---

1. How old are you?

---

2. What's your gender?

*Mark only one oval.*

- male  
 female  
 other

3. What's your knowledge in biology?

*Mark only one oval.*

	1	2	3	4	5	6	7	
none	<input type="radio"/>	expert						

4. What's your level of interest in biology?

*Mark only one oval.*

	1	2	3	4	5	
not interested	<input type="radio"/>	enthusiast				

5. What's your level of interest in games?

*Mark only one oval.*

	1	2	3	4	5	
not interested	<input type="radio"/>	enthusiast				

6. How many times have you played this video game, Hero.Coli?

*Mark only one oval.*

- never  
 first time  
 twice to five times  
 more than five times

7. Are you filling in this survey just after or just before having played?

*Mark only one oval.*

- after  
 before

### What have you understood?

---

**8. How scientifically accurate did you find the game?**

*Mark only one oval.*

- not accurate
- accurate
- I don't know

**9. In the game, which one is a genetic device?**

*(differentiate BioBrick and device)*

*Mark only one oval.*

- RBS 23%
- Hyperflagellation (low)
- I don't know
- RFP Coding Sequence

**10. Spot the non-BioBrick:**

*(differentiate BioBrick and device)*

*Mark only one oval.*

- I don't know
- Terminator
- RBS 12.6%
- Green fluorescence (medium)

**11. Spot the correct sequence:**

*Mark only one oval.*

- Promoter-Coding Sequence-RBS-Terminator
- Promoter-RBS-Coding Sequence-Terminator
- RBS-Promoter-Coding Sequence-Terminator
- I don't know

**12. Which biobrick controls only the efficiency - level of expression - of the genetic device?**

*Mark only one oval.*

- I don't know
- Promoter
- RBS
- Coding Sequence

**13. Which biobrick controls what is produced by the genetic device?**

*Mark only one oval.*

- RBS
- Promoter
- I don't know
- Coding Sequence

**14. When does GFP emit green light?**

*Mark only one oval.*

- I don't know
- When receiving blue light
- Near green rocks
- Always
- Never

**What have you understood in biology in general?**

---

Cellia (the bacterium in the game) is an E. coli. How realistic is the game compared to real-life E. coli?

The following questions refer to real-life E. coli:

15. **Real E. coli can move through the use of flagella.**

*Mark only one oval.*

- no
- yes
- I don't know

16. **Real E. coli can move by crawling like a snake.**

*Mark only one oval.*

- no
- yes
- I don't know

17. **Real E. coli have eyes.**

*Mark only one oval.*

- no
- yes
- I don't know

18. **Real E. coli look like tadpoles.**

*Mark only one oval.*

- no
- yes
- I don't know

19. **Real E. coli divide.**

*Mark only one oval.*

- no
- yes
- I don't know

20. **Real E. coli move in order to explore and harvest resources.**

*Mark only one oval.*

- no
- yes
- I don't know

21. **Real E. coli mainly evolve by absorbing foreign dna.**

*Mark only one oval.*

- no
- yes
- I don't know

22. **A specific E. coli bacterium evolves by getting stronger little by little.**

*Mark only one oval.*

- no
- yes
- I don't know

23. Real *E. coli* can be killed by some chemicals.

Mark only one oval.

- no
- yes
- I don't know

24. Real *E. coli* live in aquatic environments.

Mark only one oval.

- no
- yes
- I don't know

25. Real *E. coli* struggle against bubbles and rocks.

Mark only one oval.

- no
- yes
- I don't know

26. Real *E. coli* can glow green by producing GFP when exposed to blue light.

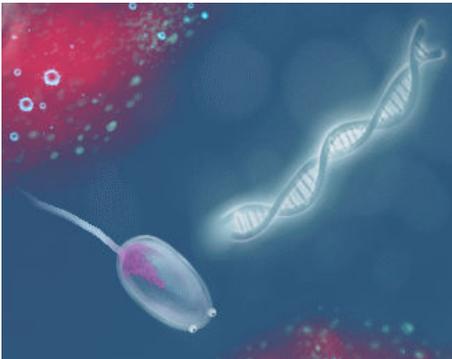
Mark only one oval.

- no
- yes
- I don't know

27. The relative sizes of DNA and real *E. coli* are accurate in the picture below.

Mark only one oval.

- no
- yes
- I don't know



## C.3 1.52 - 2017-06 to 2018-03-22

### Hero.Coli: Survey

Thanks for helping us in our project on game-based learning!

You will be asked about the game even if you haven't played yet to ensure that there are no obvious answers. Don't hesitate to use the "I don't know" option.

Let's get a quick background first.

\* Required

**1. Are you interested in video games? \***

*Mark only one oval.*

- Not at all
- Slightly
- Moderately
- A lot
- Extremely
- I don't know

**2. Do you play video games? \***

*Mark only one oval.*

- Not at all
- Rarely
- Moderately
- A lot
- Extremely
- I don't know

**3. How old are you? \***

---

**4. What is your gender? \***

*Mark only one oval.*

- Other
- Male
- Female
- Prefer not to say

### Basic biology questions

Basic biology questions

**5. How long have you studied biology? \***

*Mark only one oval.*

- Not even in middle school
- Until the end of middle school
- Until the end of high school
- Until bachelor's degree
- At least until master's degree
- I don't know

**6. Are you interested in biology? \***

*Mark only one oval.*

- Not at all
- Slightly
- Moderately
- A lot
- Extremely
- I don't know

**7. Before playing Hero.Coli, had you ever heard about synthetic biology? \***

*Mark only one oval.*

- Yes
- No
- I don't know

**8. Before playing Hero.Coli, had you ever heard about BioBricks? \***

*Mark only one oval.*

- Yes
- No
- I don't know

## Your experience with Hero.Coli

Which game versions have you played?

**9. Have you ever played an older version of Hero.Coli before? \***

Previous versions had no tutorials, no cut scenes, and a static, heavy craft interface.

*Mark only one oval.*

- Multiple times
- A few times
- Once
- No
- I don't know

**10. Have you played the current version of Hero.Coli? \***

The current version has tutorials, cut scenes, and an animated, simpler craft interface.

*Mark only one oval.*

- Multiple times
- A few times
- Once
- No
- I don't know

**11. Have you played the arcade cabinet version of Hero.Coli? \***

Hero.Coli has been demonstrated on a dedicated arcade cabinet at a few events.

*Mark only one oval.*

- Multiple times
- A few times
- Once
- No
- I don't know

**12. Have you played the Android version of Hero.Coli? \***

Hero.Coli has been demonstrated on tablets at a few events.  
*Mark only one oval.*

- Multiple times
- A few times
- Once
- No
- I don't know

## General mechanics of the game

Let's see what you understand about the game!

**13. In order to modify the abilities of the bacterium, you have to... \***

*Mark only one oval.*

- Edit the DNA of the bacterium
- Move the bacterium
- Divide the bacterium
- I don't know
- Gather nanorobots

**14. What are BioBricks and devices? \***

*Mark only one oval.*

- Proteins
- DNA sequences
- Amino-acids
- RNA sequences
- I don't know

## BioBricks

Let's see what you understand about synthetic biology!

**15. What is the name of this BioBrick? \***



*Mark only one oval.*

- Trespasser
- Transcriptor
- I don't know
- Translator
- Terminator

**16. What is the name of this BioBrick? \***



*Mark only one oval.*

- I don't know
- Precursor
- Protomer
- Procuror
- Promoter

17. What is the name of this BioBrick? \*



Mark only one oval.

- Ribosome
- Fluorescence
- I don't know
- Protein
- Coding Sequence

18. What is the name of this BioBrick? \*



Mark only one oval.

- PCR
- ATP
- I don't know
- RBS
- GFP

## BioBrick functions

Let's see what you understand about synthetic biology!

19. What does this BioBrick do? \*



Mark only one oval.

- It controls when the device is active
- It controls the level of expression, and thus how much the ability will be affected
- It shows the end of the device
- I don't know
- It controls which protein is produced, and thus which ability is affected

20. What does this BioBrick do? \*



Mark only one oval.

- I don't know
- It shows the end of the device
- It controls the level of expression, and thus how much the ability will be affected
- It controls which protein is produced, and thus which ability is affected
- It controls when the device is active

21. What does this BioBrick do? \*



Mark only one oval.

- It controls the level of expression, and thus how much the ability will be affected
- It controls when the device is active
- It controls which protein is produced, and thus which ability is affected
- I don't know
- It shows the end of the device

22. What does this BioBrick do? \*



Mark only one oval.

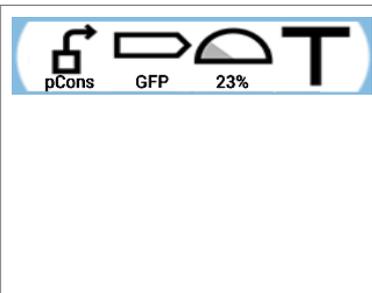
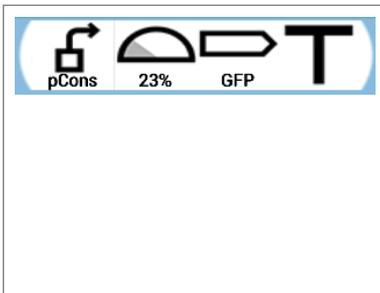
- I don't know
- It shows the end of the device
- It controls the level of expression, and thus how much the ability will be affected
- It controls which protein is produced, and thus which ability is affected
- It controls when the device is active

## Devices

Let's see what you understand about synthetic biology!

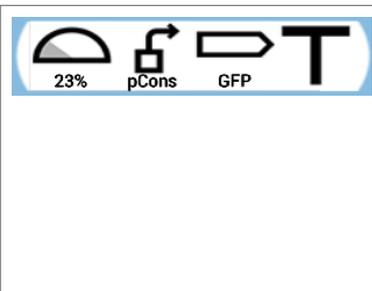
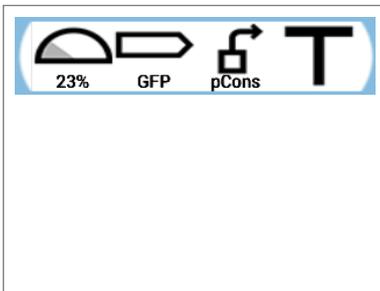
23. Pick the case where the BioBricks are well-ordered: \*

Mark only one oval.



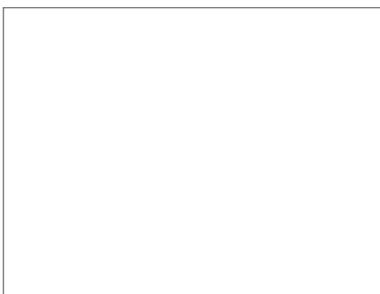
Option 1

Option 2



Option 3

Option 4



I don't know

24. When does green fluorescence happen? \*

Mark only one oval.

- Under blue light, all the time
- In front of the doors, all the time
- I don't know
- In front of the doors, when the GFP device is equipped
- Under blue light, when the GFP device is equipped

25. What happens when you unequip the movement device? \*

Mark only one oval.

- Nothing
- Flagella quickly disappear one by one
- I don't know
- The bacterium dies
- The bacterium glows

26. What is this? \*



Mark only one oval.

- I don't know
- A plasmid - it makes it possible to equip an additional device
- A nanobot - a game bonus
- Algae from the game's scenery
- An induced promoter, which works only in arabinose

## Devices symbols

Let's see what you understand about synthetic biology!

27. What does this device do? \*



Mark only one oval.

- It generates antibiotic resistance
- It makes it possible to move faster
- It generates green fluorescence in presence of l-arabinose
- It generates green fluorescence
- I don't know

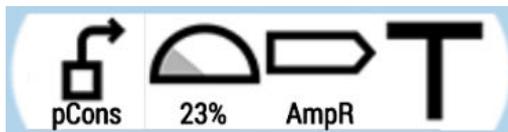
28. What does this device do? \*



Mark only one oval.

- It generates green fluorescence
- I don't know
- It generates antibiotic resistance
- It makes it possible to move faster
- It generates green fluorescence in presence of l-arabinose

29. What does this device do? \*



Mark only one oval.

- It generates green fluorescence in presence of l-arabinose
- I don't know
- It generates green fluorescence
- It makes it possible to move faster
- It generates antibiotic resistance

30. What does this device do? \*



Mark only one oval.

- It generates green fluorescence
- It generates green fluorescence in presence of l-arabinose
- It generates antibiotic resistance
- I don't know
- It makes it possible to move faster

## Devices symbols

Let's see what you understand about synthetic biology!

31. What does this device do? \*



Mark only one oval.

- It makes it possible to move faster
- I don't know
- It generates antibiotic resistance
- It generates green fluorescence in presence of l-arabinose
- It generates green fluorescence

32. What does this device do? \*



Mark only one oval.

- It generates green fluorescence in presence of l-arabinose
- It generates green fluorescence
- I don't know
- It generates antibiotic resistance
- It makes it possible to move faster

33. What does this device do? \*



Mark only one oval.

- It generates antibiotic resistance
- It makes it possible to move faster
- It generates green fluorescence in presence of l-arabinose
- I don't know
- It generates green fluorescence

34. What does this device do? \*



Mark only one oval.

- It generates green fluorescence in presence of l-arabinose
- I don't know
- It generates antibiotic resistance
- It generates green fluorescence
- It makes it possible to move faster

## Beyond the game

Less obvious aspects of the game.

35. **Guess: what would a device producing l-arabinose do, if it started with a l-arabinose-induced promoter? \***

Mark only one oval.

- After being induced, it would produce more and more l-arabinose, because it would induce itself
- It would be active only in l-arabinose clouds
- I don't know
- It would produce nothing since it induces itself
- It would produce l-arabinose all the time

36. **Guess: the bacterium would glow yellow... \***

Mark only one oval.

- If it produced BFP under purple light
- If it produced GFP under yellow light
- If it produced RFP under yellow light
- If it produced YFP under cyan light
- I don't know

37. What is the species of the bacterium of the game? \*

Mark only one oval.

- I don't know
- Hero.Coli
- E. Coli
- Cella
- Nanobot

38. What is the scientific name of the tails of the bacterium? \*

Mark only one oval.

- I don't know
- Flagella
- Fins
- Plasmids
- Mitochondria

39. Find the antibiotic: \*

Mark only one oval.

- GFP
- Arabinose
- Ampicillin
- I don't know
- Terminator

## Remarks

You're almost done!

40. You can write down remarks here.

---

---

---

---

---

## Thanks for filling in this survey!

This section's role is to associate your answers in this questionnaire with this pre-filled Hero.Coli anonymous ID. Please do not edit it.

41. Do not edit - pre-filled anonymous ID \*

---

# C.4 1.52.2 / 1.60 / 1.61 - 2018-03-23 onwards

## Your experience with Hero.Coli

Thank you for participating in this anonymous survey! (10 questions, 5 min)

This video game has been developed at the CRI - Paris-Descartes University in Paris, France. Its purpose is to have you discover synthetic biology, and to do research. This research needs your input in this survey.

Gathered data are anonymous and open. Contact us for more information: [herocoli.game@gmail.com](mailto:herocoli.game@gmail.com)

\* Required

### 1. Are you interested in learning more about... \*

Mark only one oval per row.

	Not at all	Slightly	Moderately	A lot	Extremely
Biology	<input type="radio"/>				
Synthetic biology	<input type="radio"/>				
Video games	<input type="radio"/>				
Engineering	<input type="radio"/>				

### 2. Have you ever played Hero.Coli? \*

Mark only one oval.

- I just played for the first time *Skip to question 10.*
- I played it multiple times recently on this computer *Skip to question 10.*
- I played recently on an other computer
- I played it some time ago
- No / not yet

## Your profile

Only these 7 questions are required to unlock the game. You will be given a chance to answer more questions and make a greater contribution to our research project.

### 3. How old are you? \*

Enter a number, without "years old"

### 4. \*

Mark only one oval per row.

	Female	Male	Other	Prefer not to say
What is your gender?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### 5. \*

Mark only one oval per row.

	Not at all	Slightly	Moderately	A lot	Extremely
Are you interested in video games?	<input type="radio"/>				
Are you interested in biology?	<input type="radio"/>				

### 6. \*

Mark only one oval per row.

	Not even in middle school	Until the end of middle school	Until the end of high school	Until bachelor's degree	At least until master's degree
How long have you studied biology?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7. \*

Mark only one oval per row.

	Not at all	Rarely	Moderately	A lot	Extremely
Do you play video games?	<input type="radio"/>				

8. \*

Mark only one oval per row.

	No	Yes, but I don't exactly know what it means	Yes, and I know what it means
Have you ever heard about synthetic biology or BioBricks, outside of Hero.Coli?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Volunteering

9. Do you volunteer to contribute to our study by answering 9 more questions? (5 min) \*

Your contribution would be greatly appreciated. You don't need to answer these additional questions to unlock the game though: you are finished if you answer "No".

Mark only one oval.

- Yes  
 No Skip to question 19.

You will be asked about the game even if you haven't played yet to ensure that there are no obvious answers. Don't hesitate to use the "I don't know" option.

10. \*

Mark only one oval per row.

	Not applicable: not played yet	Not at all	A bit	Moderately	A lot	Extremely
Did you enjoy playing the game?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11. \*

Mark only one oval per row.

	Gather nanorobots	Edit the DNA of the bacterium	Move the bacterium	Divide the bacterium	I don't know
In order to modify the abilities of the bacterium, you have to...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

12. \*

Mark only one oval per row.

	RNA sequences	Amino-acids	Proteins	DNA sequences	I don't know
What are BioBricks and devices?	<input type="radio"/>				

13. \*

Mark only one oval per row.

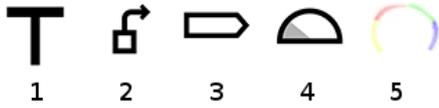
	Ampicillin	Arabinose	GFP	Terminator	I don't know
Find the antibiotic:	<input type="radio"/>				

## BioBricks

Let's see what you understand about synthetic biology!

**14. BioBricks: \***

There can be 0, 1 or more correct answers.



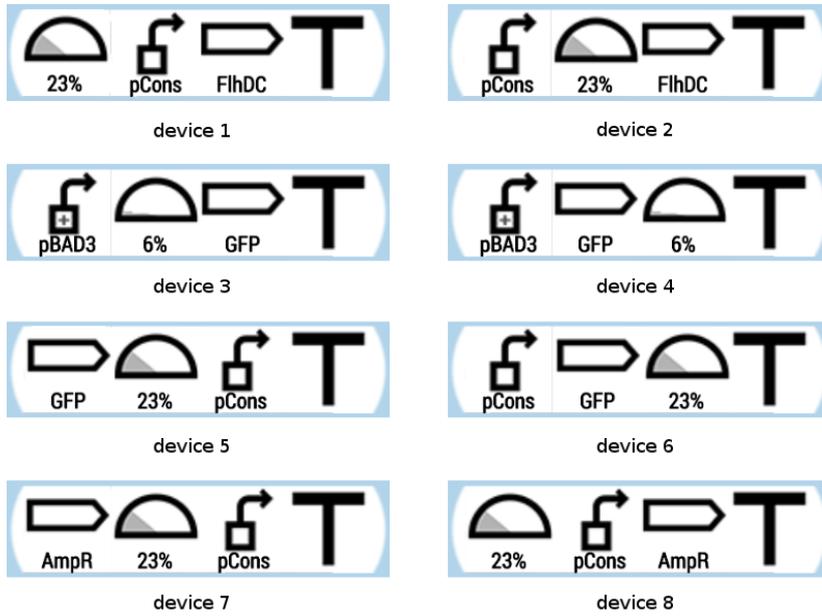
Mark only one oval per row.

	None of these	1	2	3	4	5	I don't know
Plasmid is...	<input type="radio"/>						
Represents the end of a device...	<input type="radio"/>						
Promoter is...	<input type="radio"/>						
Represents the ability given...	<input type="radio"/>						
Terminator is...	<input type="radio"/>						
Codes a protein...	<input type="radio"/>						
RBS is...	<input type="radio"/>						
Can represent GFP...	<input type="radio"/>						
Coding Sequence is...	<input type="radio"/>						
Controls when the device is active...	<input type="radio"/>						
Controls the level of expression, and thus how much the ability will be affected...	<input type="radio"/>						
Makes it possible to equip an additional device.	<input type="radio"/>						
Operator is...	<input type="radio"/>						

**Devices**

Let's see what you understand about synthetic biology!

15. \*



Mark only one oval per row.

	The bricks are not well-ordered	It generates green fluorescence	It generates green fluorescence in presence of arabinose inducer	It makes it possible to move faster	It generates antibiotic resistance	I don't know
device 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
device 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
device 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
device 4	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
device 5	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
device 6	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
device 7	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
device 8	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

16. \*

Mark only one oval per row.

	Under blue light, when the GFP device is equipped	Under blue light, all the time	In front of the doors, when the GFP device is equipped	In front of the doors, all the time	I don't know
When does green fluorescence happen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

17. \*

Mark only one oval per row.

	The bacterium glows	The bacterium dies	Flagella quickly disappear one by one	Nothing	I don't know
What happens when you unequip the movement device?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**18. Last question. Next page only contains remarks. \***

*Mark only one oval per row.*

	It produces arabinose all the time	It is active only in arabinose clouds	It produces more and more arabinose after being induced, because it induces itself	It produces nothing since it induces itself	I don't know
Guess: you have crafted a functional device containing an arabinose-induced promoter and an arabinose Coding Sequence (CDS). What will happen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Thanks for filling in this survey!**

You're done!

**19. You can write down remarks here.**

---

---

---

---

---

**20. ⚠ Do not edit \***

This section's role is to associate your answers in this questionnaire with this pre-filled Hero.Coli anonymous ID. Please do not edit it. If you accidentally erased it, type CTRL + Z on Windows and Linux, or CMD (⌘) + Z on Mac, until the ID reappears.

---