Unfolding the dynamics of creativity, novelties and innovation

White paper of the Kreyon project

Vittorio Loreto
Sapienza University of Rome,
Physics Department, Rome, Italy
ISI Foundation, Turin, Italy
SONY-Computer Science Lab, Paris, France
(Dated: March the 20th 2015)

The Kreyon project addresses the dynamics of novelties - a fundamental factor in the evolution of human societies, biological systems and technology- with the aim to unfold and quantify the underlying mechanisms through which creativity emerges and innovations diffuse, compete and stabilize. The project is timely due to the availability of extensive longitudinal records of human, social and technological evolution. Kreyon will exploit the unique opportunity offered by the combination of ICT tools for social computation with powerful analytical and modeling tools, by blending, in a unitary interdisciplinary effort, three main activities: web-based experiments, data science and theoretical modeling.

INTRODUCTION

Novelty is part of our everyday experience more than we usually think. We continuously meet new people, adopt new words, listen to new songs, watch a new movie or a video, use a new technology. In addition we constantly leverage our creativity as a natural mode of human nature of finding new valuable solutions to old and emerging problems [1]. At a collective level, innovation is a fundamental factor in the evolution of biological systems [2–5], human societies [6–8] and technology [9–14]. And of course creativity and innovation are key for scientific discoveries [15]. It is interesting to report here a passage from the seminal paper of Herbert Simon [16]:

Many scientists and philosophers have doubted whether a scientific explanation of discovery is possible or whether there is even something describable as a method of discovery. Einstein himself is quoted as saying, "There is no logical path leading to [scientific] laws. They can only be reached by intuition, based upon something like an intellectual love [Einfuehlung] of the objects of experience." If Einstein meant by "logical path" a deductive route to discovery, he was surely right. If he meant to say that there is not method in the madness of discovery, we might question-and investigate empirically-whether his pessimism was justified.

Studies on creativity have a longstanding history and a comprehensive account of it is beyond the scope here. In particular scholars have long debated about a suitable definition of creativity [17–19] as well as about ways to measure it [20–23]. Many studies focus on identifying the neural correlates of creativity [24–26]. In addition a huge literature exists on different aspects of innovation, concerning both its adoption and diffusion [7, 27–30], as well as the creative processes through which it is generated, e.g., processes like recombination [9, 13], tinkering [2] or exaptation [3, 31]. Major innovations have been studied in cultural [8], technological [14], and biological systems [2, 4, 32]. Still [23, 33] creativity and innovation have proved to be resistant to measurement, understanding and control.

Following Ken Robinson [34], creativity can be thought of as the *process of developing original ideas that have value*, while innovation as the *process of putting new ideas into practice*. From this perspective creativity and innovation are seen as applied imagination, and as such, they are becoming progressively more and more crucial to help us, as individuals and societies, to face the challenges of our complex world [33]. The largest fraction of the mankind, in fact, will soon adopt technologies that have yet to be invented or have jobs that do not exist.

THE OPPORTUNITY

Our societies are being thoroughly transformed by the pervasive role technology is playing on our culture and everyday life. Techno-social systems is the locution more and more adopted to quickly refer to social systems in which technology entangles, in an original and unpredictable way, cognitive, behavioral and social aspects of human beings. Technology helps connecting people and circulating information, and affects more and more the way humans interact with each other. Everyday, a huge amount of information is exchanged by people through posts and comments on-line, tweets or emails, or phone calls as a natural aptitude of humans to share news, thoughts, feelings, or experiences. This revolution does not come without a cost and in our complex world always new global challenges emerge that call for new paradigms and original thinking to be faced: climate change, global financial crises, global pandemics, growth of cities, urbanisation and migration patterns. In this framework we progressively face the need to increase the number of people able to imagine original and valuable solutions [35] to sustain large human societies safely and prosperously.

Creativity and innovation are key elements in many different areas and disciplines since they represent the first motor to explore new solutions in ever-changing and unpredictable environments. New biological traits and functions, new technological artefacts, new social, linguistic and cultural structures, new meanings, are very often triggered by the mutated external conditions. Unfortunately the detailed mechanisms through which humans, societies and nature express their creativity and innovate are largely unknown and no mathematical framework have been satisfactorily proposed so far. In addition, creativity and innovation have long been seen as pertaining to different realms and as such treated independently. Creativity has often been seen as the outcome of a sudden individual discontinuity and, as such, ineffable and out of reach for a scientific approach. Studies on innovation, on the other hand, have mainly been focused on how novelties can diffuse and be adopted at a large scale as well as on the specific pathways through which creativity can emerge: e.g., recombination, exaptation, tinkering, serendipity, trial and error. Creativity is nowadays seen more and more also as a social process affected by the interactions individuals experience, the pieces of knowledge they are exposed to, the challenges they have to face and the constraints they experience. In this sense one speaks more of "Cultures of Creativity" [36]. Still, creativity and innovation have proved to be resistant to measurement, understanding and control.

The Kreyon project is proposing for the first time a unitary approach where creativity and innovation are seen as two sides of the same coin and jointly investigated and modeled. In doing this it exploits the extraordinary alignment of three circumstances:

(i) The ability to monitor, quantify and model human behaviors at unprecedented levels of resolution and scale, unleashed by the planetary-scale adoption of the World Wide Web, mobile communication technologies, e-commerce systems, and on-line social networks. The possibility to access to digital fingerprints of individuals is opening tremendous avenues for an unprecedented monitoring at a "microscopic level" of collective phenomena involving human beings. We are moving very fast towards a sort of tomography of our societies, with a key contribution of people acting as data gathering "sensors".

(ii) The opportunities web-gaming and social computation are offering to the emergence of new forms of participation arising from the interplay of ICT services and communities of citizens. In the last few years the Web has been progressively acquiring the status of an infrastructure for social computing that allows researchers to coordinate the cognitive abilities of users in online communities, and to suggest how to steer the collective action towards predefined goals. This general trend is also triggering the adoption of web-games as a very interesting laboratory to run experiments in the social-sciences and whenever the peculiar human computation abilities are crucially required for research purposes.

(iii) The maturity of complex systems and data science applied to socio-technical systems. The theoretical and modeling tools recently developed by physicists, mathematicians, computer and social scientists to analyse, interpret and visualize complex data sets have reached the maturity to effectively address the challenges of our era. The concurrence of all these elements is opening tremendous opportunities towards an understanding of the complexity of our societies, with the final goal of deploying human imagination for the betterment of our communities and even civilization. This project aims at timely leveraging these possibilities to investigate creativity and innovation processes in a quantitative way by blending in a unique ambitious project data-science, web-based experiments and theoretical modeling.

Building on these three pillars, the project promises to construct a quantitative framework through which important questions could be raised, refined, polished and eventually answered. We expect, in this way, to trigger a revamped interest of the scientific community and the society at large towards a systematic and quantitative approach to creativity and innovation. The needed renewed effort of putting again creativity and innovation under the lens of scientific investigation, does not mean having the presumption of being able to predict the when/where/what of a creative act [37], rather that of understanding and controlling the process and the environment leading to something innovative.

NOVELTIES, INNOVATION AND THE ADJACENT POSSIBLE

Creative solutions, novelties and innovation share an important feature: often innovative events do not happen by chance, rather they seem to be triggered by some previous novelty or innovation. In studies of biological, technological, and cultural evolution, it has been hypothesized that one innovation can lay the groundwork for another by creating fresh opportunities. In our daily lives, a similar process may account for why one new thing so often leads to another. This idea has been beautifully summarized by the notion of *adjacent possible* introduced by Stuart Kauffman [4, 38]. In this picture the advance into the adjacent possible is the driving force for correlating innovative events, and novelties are produced through an exploration of a space - physical, conceptual, technological or biological - that enlarges itself whenever one reaches a point of the space never touched before. Originally introduced in the framework of biology, the adjacent possible metaphor already expanded its scope to include all those things (ideas, linguistic structures, concepts, molecules, genomes, technological artefacts, etc.) that are one step away from what actually exists, and hence can arise from incremental modifications and recombination of existing material. In Steven Johnson's words: *The strange and beautiful truth about the adja-*

cent possible is that its boundaries grow as one explores them.

FIG. 1. Metaphor of the adjacent possible in terms of doors opening on new possible doors *ad libitum*.

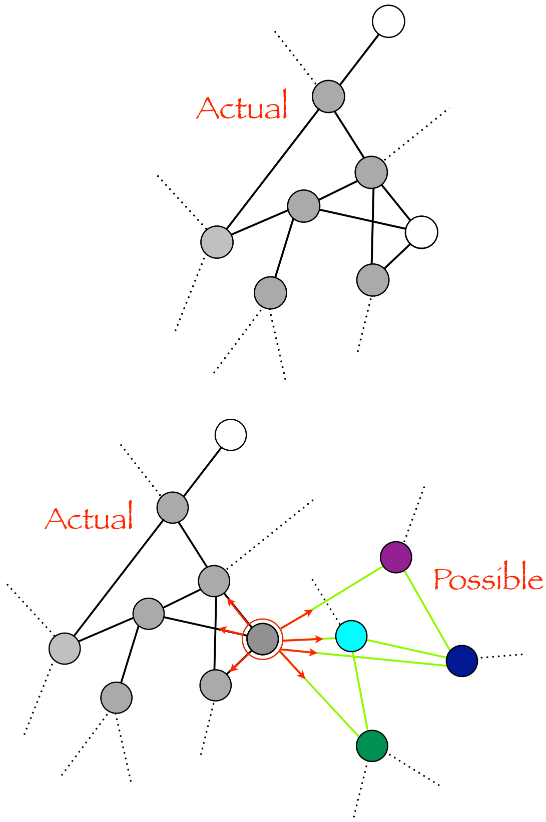

FIG. 2. Mathematical illustration of the adjacent possible in terms of a graph that conditionally expands whenever a walker visits for the first time a node.

The very definition of adjacent possible encodes the dichotomy between the *actual* and the *possible* [39]: the actual realisation of a given phenomenon and the space of possibilities still unexplored. Fig. 2 illustrates with a cartoon this idea. A walker is wandering on the nodes of a graph. The grey nodes are those already visited in the past while the white ones are the ones never visited. Once the walker visits for the first time a white node another part of the graph appears that could not even be foreseen before visiting that node.

Though the creative power of the expansion into the adjacent possible is widely appreciated at an anecdotal level, still its importance in the scientific literature [14, 38, 40–43] is, in my opinion, underestimated. As a consequence the whole idea remains poorly understood theoretically and undocumented empirically. Recently, in collaboration with Vito Servedio, Steven Strogatz and Francesca Tria, I have introduced [44] a first mathematical model of the dynamics of novelties correlated via the adjacent possible, and derive three testable, quantitative predictions from it. Those predictions were shown to be borne out in several data sets drawn from social and technological systems. The model we introduced is a generalization of Polya's urn [45–47]. In the classical version of this model [45], balls of various colors are placed in an urn. A ball is withdrawn at random, inspected, and placed back in the urn along with a certain number of new balls of the same color, thereby increasing that color's likelihood of being drawn again in later rounds. The resulting "rich-get-richer" dynamics leads to skewed distributions [48, 49] and has been used to model the emergence of power laws and related heavy-tailed phenomena in fields ranging from genetics and epidemiology to linguistics and computer science [50–52].

In our generalisation a urn $\mathcal{U}$, initially containing N_0 balls of distinct colors, represents the set of possibilities, i.e., the adjacent possible. The actual is here represented by the series, $\mathcal{S}$, of the extractions drawn from the urn. Just as the adjacent possible expands when something novel occurs, the contents of the urn itself are assumed to enlarge whenever a novel (never extracted before) element is withdrawn. The evolution proceeds through two main mechanisms: *reinforcement* and *expansion of the adjacent possible*. At each time step t one selects an element s_t at random from $\mathcal{U}$ and records it in the sequence. We then put the element s_t back into $\mathcal{U}$ along with ρ additional copies of itself. The parameter ρ represents a *reinforcement* process, i.e., the more likely use of an element in a given context. The key assumption concerns what happens if (and only if) the chosen element s_t happens to be novel (i.e., it is appearing for the first time in the sequence $\mathcal{S}$). In that case we put $\nu + 1$ brand new and distinct elements in the urn. These new elements represent the set of new possibilities triggered by the novelty s_t. Hence $\nu+1$ is the size of the new adjacent possible made available once we have a novel experience. This *expansion of the adjacent possible*, conditioned on the occurrence of a novelty, is the crucial ingredient in this modeling scheme. Fig. 3 illustrates these dynamical rules and I refer to [44] for a full account of the whole phenomenology.

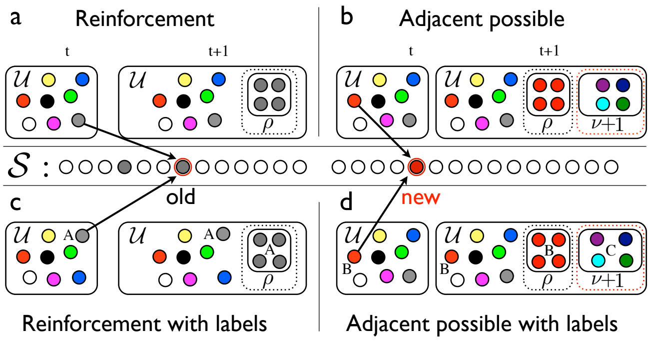

FIG. 3. Simple urn model with triggering **(a,b)** and urn model with semantic triggering **(c,d)**. **(a)** Generic reinforcement step of the evolution. An element (the gray ball) that had previously been drawn from the urn $\mathcal{U}$ is drawn again. In this case one adds this element to $\mathcal{S}$ (depicted at the center of the figure) and, at the same time, puts ρ additional gray balls into $\mathcal{U}$. **(b)** Generic adjacent possible step of the evolution. Here, upon drawing a new ball (red) from $\mathcal{U}$, $\nu+1$ brand new balls are added to $\mathcal{U}$ along with the ρ red balls of the reinforcement step that takes place at each time step. **(c,d)** Urn model with semantic triggering. Same as above except that now each ball has a label defining its semantic context. The label is conserved during a reinforcement event (e.g., the label A for the gray balls on panel c) while it appears as a brand new label, C, for the $\nu+1$ balls added for an adjacent possible event (panel d).

The above model predicts statistical laws for the rate at which novelties happen (Heaps' law) and for the probability distribution on the space explored (Zipf's law), as well as signatures of the process by which one novelty sets the stage for another. The predictions of this models have been tested on four data sets of human activity: the edit events of Wikipedia pages, the emergence of tags in annotation systems, the sequence of words in texts, and listening to new songs in online music catalogues. By providing the first quantitative characterization of the dynamics of correlated novelties, these results provide a starting point for a deeper understanding of the adjacent possible and the different nature of triggering events (timeliness, spreading, individual vs. collective properties) that are likely to be important in the investigation of biological, linguistic, cultural, and technological evolution. The Kreyon projects starts from this very preliminary point to setup a mathematical, computational and experimental approach to creativity and innovation.

KREYON'S OBJECTIVES

Deeply rooted in the above mentioned context, the Kreyon project aims at raising and addressing (without being limited to) the following scientific questions:

a) Signatures of the *adjacent possible* at the individual and collective level. What is its structure and how does it get dynamically reshaped by individual innovative events?

b) Operational definition of creativity. To which extent creativity is an individual or a social process? How to quantify the value of a new idea? Great ideas could be useless. But some of them could have a value, though not immediately evident. And also, does creativity require (or is enhanced by) specific constraints?

c) Microscopic mechanisms at play. Can we quantify the role of tinkering, recombination, serendipity, trial/error?

d) Understanding the deep links among individual creativity, the emergence of novelties and the collective phenomena in innovation dynamics. Novelties should not necessarily be new to the whole history. Some could be new to someone, his own context, workplace and only some of them new to the whole history. How can one tell them apart?

e) A matter of scales. Do creativity and innovation proceed gradually or through large leaps [53]? Individual vs. collective effects for innovation. Some new idea could be adjacent for someone and very far for the community (”That's one small step for [a] man, one giant leap for mankind” effect).

f) Paths leading from the early adoption to a large-scale spreading of innovations. How several innovations compete, fail and get successful? Why sometimes innovations can fail because too far ahead of their times?

g) Correlation among creative and innovative events. Can one find signatures and of the complex triggering processes among innovations, as predicted by the notion of adjacent possible?

h) Adjacent possible in action. Is it possible to identify the best environments and the best strategies to foster creativity and innovation?

DESCRIPTION OF THE ACTIVITIES

In order to address the above mentioned key questions the project is structured in three main Workpackages (WP) whose continuous cross-fertilization will be managed to produce a virtuous loop linking empirical results, hypotheses, theories and testing.

WP1: Data science approach to creativity and innovation

The aim of this Workpackage is to look for patterns of creativity and signatures of correlations, precursors and triggering in several areas of social dynamics, cognitive sciences and technological evolution. To this end large relevant datasets will be gathered, by leveraging the availability of extensive longitudinal records of human activities online, and adapted to the specific needs. A key feature to be collected will be the time-stamp of individual events in order to reconstruct ordered time-series in the different contexts. This will be useful to investigate triggering processes, correlations and cause-effect relationships.

While analyzing the different databases mentioned we shall be using several tools to extract, aggregate and make sense of the huge amount of data gathered. Our analysis will be aimed at:

- revealing general statistical laws for the rate at which novelties happen (e.g., Heaps' law) and for the probability distribution on the space explored (e.g., Zipf's law).
- detecting triggering events and correlations among semantically related events. For instance, in the case of music, one can imagine that when we first discover an artist or a composer that we like, we might want to learn more about his or her work. This in turn can stimulate us to listen to other songs by the same artist. To this end we shall devise and adopt suitable measures (for instance entropic measures) to tell apart true signature of correlations from purely spurious statistical effects. Suitable null-models will be devised on a case-by-case basis. We shall also compare triggering events at the individual (single books, single users, etc.) and collective (aggregate) level with the aim of investigating whether the statistical signatures one detects emerge as the outcome of a collective process or are present also at the level of the single user.
- measuring the level of originality and novelty of a specific piece of knowledge with respect to the previous production in its specific area. For instance how new is a text or a song with respect to whatever preceded them? In some cases suitable sets of metadata could be used to this end. More in general it will be important using tools to define a notion of distance between two pieces of knowledge, e.g. two sequences of characters. In nature many systems and phenomena are often represented in terms of sequences or strings of characters. In this area we plan to adopt data-compression oriented approaches [54, 55] to define a notion of remoteness and distance between pairs of sequences or strings of characters on the basis of their relative information content (relative entropy).
- moving towards an operational definition of creativity as encoded in data structures mirroring human activities. We shall exploit all the metadata attached to pieces of knowledge as well their complex network structure to extract signatures of creative events or actions, long-term impact, centrality measures, quantitative evaluation of influences (see e.g., the recent paper [**?**].
- style detection and its evolution. A natural by-product of the information-theoretic and the complex networks oriented approaches discussed above, concerns the possibility to automatically extract relevant statistical features of a piece of knowledge, i.e., its stylistic features. This in turn allows to identify where specific stylistic traits come from, how did they get recombined (recombining in a *different* way old materials) and modified over time. This represents an interesting opportunity since it will represent a proxy to trace back the origin and evolution of specific creative and innovative events.
- detecting small signals. When looking for signatures of innovative events an interesting problem concerns whether the novelty reveals itself in a statistically significant way or rather appears as an isolated pattern (statistically irrelevant). We plan to develop specific tools to detect relevant and meaningful, though not frequent, events.
- visualisation tools. Finally Kreyon plans to adopt the whole machinery of graph and network theory map, analyze and visualize complex relationships among innovative events.

Our analysis will be focuses (though not limited to) on the following environments, each characterized by specific datasets.

a. Creativity and innovation in language Language and textual corpora represent a valuable source to explore how novelties emerge, get adopted, have success and eventually die. Examples of the datasets we shall consider include: (a) the Gutenberg corpus of free ebooks [1] including roughly 100,000 digitized and diligently proofread ebooks; (b) Google Books Ngrams [56][2] coming from the huge Google's effort to digitize a good deal of the books ever published; (c) the Wikipedia dataset [3], including the description of the individual events, creation of a new page or edit of an old one, with details on the time stamp, the user, the unique IDs of all the pages ad their complex network of connections; (d) the Corpus Of Historical American english (COHA)[4]; (e) tags coming from social annotation systems, e.g. delicious[5].

[1] http://www.gutenberg.org/
[2] https://books.google.com/ngrams
[3] http://www.wikipedia.org/
[4] http://corpus.byu.edu/coha/
[5] e.g.: http://delicious.com/

b. Creativity and innovation in music Datasets about music are another important source of information concerning creativity, novelties and innovations. Data from Last.fm [6], for instance, a music website equipped with a music recommender system, allows to access detailed profiles of each user's musical taste by recording details of the songs the user listens to, either from Internet radio stations, or the user's computer or many portable music devices. Through Last.fm we shall be able to access the listening histories of individuals and follow the way in which users experience novelties and how they explore their adjacent possible. In addition detailed data about each track can be obtained allowing to explore the success of individual tracks as well as the chains of influences. On the other hand the Million Song Dataset [7][57] contains audio features and metadata for a million contemporary popular music tracks. This represents an invaluable source of information on triggering events, emergence of genre, trends, etc. By meshing up all the available data, we plan to be able to trace the whole process going from the introduction of a novelty, its adoption both at the individual and the collective level and its overall history.

c. Creativity and innovation in art Art is perhaps the domain of human activities with the strongest association with creativity. Several datasets are presently available through, for instance, the Google Cultural Institute [8], ArtCyclopedia [9], the OCAD University (the university of the imagination) database [10] or the wikimedia commons for art [11]. We plan to use them, along as the suite of metadata they come with, to trace also in this area the processes underlying the artistic act, their recombination, diffusion and their small or large leaps in the space of artistic paradigms.

d. Creativity and innovation in science Scientific production reveals a lot about the way in which discoveries follow each other, affect each other, reflect the collective knowledge of their era or simply stand out as seminal discontinuities. Databases as arXiv.org reveal an extremely rich picture of the dynamics of the scientific activity, by monitoring who contributed which results, in collaboration with whom, citing which previous papers. All this amounts to roughly one million e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics, each with their own time-stamp, complete history and metadata. On a longer timescale the American Physical Society (APS) [12] provides data for all papers published on its journals since 1893, with the added value, for each paper, of the so-called PACS (Physics and Astronomy Classification Scheme) along with the citation networks: who cited whom and when. We plan to use the ensemble of these information to map the structure and the evolution of the scientific adjacent possible, both at the individual and at the level of the different scientific communities.

e. Creativity and innovation in technology Data about patents can reveal a great deal about the way in which new ideas are produced, recombined, merged and eventually exploited. The National Bureau of Economic Research [13], for instance, comprise detail information on almost 3 million U.S. patents granted between January 1963 and December 1999, all citations made to these patents between 1975 and 1999 (over 16 million), and a reasonably broad match of patents to Compustat (the data set of all firms traded in the U.S. stock market).

WP2: Game-based experiments

The human ability to solve those tasks traditionally difficult for computers has been largely exploited in fields as image labeling (e.g., the collaborative ESP game [58]), pattern recognition (e.g., reCAPTCHA), etc. The use of games and more specifically web-based games [59] for research purposes is a fast spreading phenomenon, changing the way research activities are conduced and how data are generated in many scientific fields [60]. The Kreyon team has recently introduced a novel general purpose web-based platform for social computation [61], Experimental Tribe (XTribe, www.xtribe.eu, still in beta version), a platform to run focused web-based experiments aiming at elucidating basic mechanisms at play whenever we learn, create and innovate.

This WP will tap on the XTribe experience to run suitably devised games aimed at elucidating how creativity shows up in controlled experimental tasks. Playing has a natural link with learning. Play is the fundamental context for most human learning across ages: activities like exploration, experimentation, innovation, testing the limits, etc. all happens best when playing. In single or multiple players games, users will be asked to come up with a non-trivial solution for a specific constrained task. The game set-up will allow to fully monitor how people explore, discover and shape their own space of possible solutions and how their strategies are affected by the framework they are asked to play in: for instance limited time availability, number and quality of the constraints imposed; level of abstraction of the task; number of players, etc.

For concreteness let us briefly illustrate a few potential games Kreyon will focus on.

- Taboo-like games. In this kind of games two or more players play asymmetric games, a *Hearer* has to cor-

[6] http://last.fm
[7] http://labrosa.ee.columbia.edu/millionsong/
[8] https://www.google.com/culturalinstitute/project/art-project
[9] http://www.artcyclopedia.com/
[10] http://ocad.libguides.com/content.php?pid=229447&sid=2698044
[11] http://commons.wikimedia.org/wiki/Category:Art
[12] publish.aps.org
[13] http://www.nber.org/patents/

rectly guess a concept/meaning a *Speaker* tries to spell out in words, avoiding a certain number of *taboo*-words. Taboos are implementing constraints forcing users to find "alternatives", presumably creative, solutions (here creativity displays itself in finding new valuable solutions to a specific problem). Validation here is intrinsic because the hearer has to correctly guess in order for the solution to be valid. The games feature several level of abstraction for the meaning to be guessed and the gaming framework can be manipulated changing for instance the available time or the composition of the taboo list (fixed, growing, self-consistent), the number and quality of players. Variants could be implemented where meanings could be numbers (taboos being in this case numbers or numerals), or other pieces of knowledge (e.g., music tracks, paintings, etc.). More in general these games are examples of language games [62] (or, following Wittgenstein [63], communication acts) intended to convey a generic meaning with an effort to both for Speaker (coding) and Hearer (decoding). The strategies adopted by both Speaker and Hearer can be nicely investigated using the gaming framework.

- Tangram-like games. Also in this case we plan an asymmetric guessing task. Unlike the Taboo-like games the *Speaker* has to come up with a solution using exclusively shapes, for instance LEGO bricks, i.e., arranging them in a specific way that could convey the intended meaning. The *Hearer* has to guess the intended meaning. The games could also be played in an asynchronous way by recording Speaker solution and submitting it later to one or more players. The *value* of the solution will be assessed both by the level of success of the interaction and by explicit voting procedures. Also here constraints can be manipulated to play with different kinds of meanings, number of players, number and quality of the available shapes, available time, etc.music
- Copystree. In this single player game, users are presented with texts and their task is to *copy* them, very much as middle-ages copyists. In doing this each copyist is populating a tree and the experimenter monitor the whole process of creation of a tradition (in biology a phylogeny). The Task is not so *innocent* since the text to be copied is degraded in different ways, so that it cannot be read in full. Users have thus to fill the gaps and modify the texts according to their specific taste and sensibility in order to reconstruct the original text. The outcome is the production of variants whose evolution, diffusion and adoption can be fully traced back. Creativity here displays in the attempt to guess the original form by proposing original though sound versions. Variants include the possibility to access two or more texts in order to foster recombination processes (horizontal transmission in biology, borrowings in linguistics).
- Incipit-like games. In this kind of games a user will be prompted with an incipit, a piece of text, a piece of music, and her task will be that of continuing it by providing an equally long piece of knowledge consistent with the incipit. Also in this case validation will be intrinsic to the game because all the compositions will be publicly available and other users or the public at large will be able to vote, like and evaluate them. At the same time the style of compositions of subsets of them could be copied or modified by other users in a never-ending process of innovation, mutation, copy and adaptation. On the scientific side we'll have the opportunity to monitor the whole process in all its tiny details and we'll be able to extract relevant information about the most successful and popular strategies along as the dynamics of diffusion and adoptions of novelties. We shall start with the linguistic and musical domains but we shall progressively consider many other domains where creativity plays a crucial role, e.g., visual arts, photography, etc.
- Art games. Several games can be realized using artistic material (exploiting the datasets mentioned above). For instance, games will be implemented where users are requested to state the two most similar paintings among a given set of them or find the intruder. This will allow to exploit human cognitive abilities to map a network or artworks, from which extracting innovations, trends, triggering phenomena (who affected whom), recombination events (identifying borrowing or reworked patterns) and more in general to trace the history of creativity in art as seen by users. This precious datasets will be the basis to devise and run further sets of experiments. Variants could include the collaborative identification of innovative patterns or the possibility to map the adjacent possible of individual artists and communities.

WP3: Mathematical modeling of creativity and innovation

This Workpackage will be devoted to modeling schemes, mimicking the exploration and the reshaping of the adjacent possible. In [44] we introduced a mathematical framework to investigate the processes through which creativity and innovations proceeds both at the individual and collective level. We plan to explore in particular two main frameworks to investigate the dynamics and evolution of the adjacent possible. One is that of Polya's urn (PU) models [47]. This class of models is particularly suitable to our problem since it considers two spaces evolving in parallel: we can think of the urn as the space of possibilities, while the sequence of balls that are withdrawn is the history that is actually realized. The interesting variant we proposed is that the space of possibilities (i.e., the urn) grows *conditionally* to the actual realization of a novelty. This translates Kauffman's idea of the expansion in the adjacent possible. Another interesting formulation of the same problem is to look at the whole process leading to novelties and innovation as the path of a random walker wandering on a growing graph (GG), each node of the graph representing a specific state reached during the evolution. In this picture the graph structure is self-consistently reshaped whenever a novelty occurs. This basic modeling structure will represent the basis for our theoretical investigations and several specific directions will be pursued.

- Detailed investigations of the PU and GG-like modeling schemes as a function of the different parameters and graph structures. The investigation will adopt, without being limited to, agent-based models, mean-field theories, stochastic models, graph-based models. An explicit analytical understanding, paralleling the numerical investigations, will be pursued.
- Role of correlations in Innovation Dynamics. Correlations are crucially relevant since they mirror the existence of semantic links between different creative and innovative events. This is the way through which we shall investigate how *one thing leads to another* in our modeling schemes, i.e., the triggering events through which one innovation could set the stage for further innovative events.
- Role of the expansion of the adjacent possible. Nobody knows how precisely model the adjacent possible and how it gets dynamically reshaped. Our aim here is that of exploring several structures for it and several ways in which it could possibly expand. Specific predictions will be every time formulated to be grounded with actual data.
- Role of individual vs. collective factors. Here we shall explore how a community of individuals explores and reshapes collectively the adjacent possible, allowing for strong differences and continuous interactions among individuals. This line of investigation will aim at elucidating whether innovations proceed mainly gradually or through large leaps by constantly looking at what happens at the level of the single individual with respect to the whole population. The individual vs. collective level also allows to investigate the processes through which good ideas can emerge, spread, compete with others and eventually get successful or die.
- Role of time-scales in Innovation Dynamics. This line of activity will investigate the interplay among all the time-scales involved in creative acts and innovation events. In particular we shall try to elucidate what are the relevant time-scales, both at the individual and collective level, and how they are affected by the different scenarios implemented. We shall address questions concerning whether is it reasonable to observe the emergence of the same innovation in apparently disconnected communities or rather this is mainly due to underlying, though difficult to detect, correlations. At the same time we shall address the question of why certain innovation seem to be *too far ahead of their times*, i.e. the long time scale between their introduction and their success.

ENDURING IMPACT

I expect this project will stimulate a new wave of research and activities on creativity and innovation processes as key factors in many different domains. I expect a significant impulse towards quantitative and systematic studies aimed at clarifying the role of the determinants of creativity. The notion of adjacent possible could play a major role in shaping the research agendas in many different domains such as physics, mathematics, biology, computer science along as psychology, cognitive science, linguistics, sociology and economics. As a consequence I expect that sectors like education, learning, technological evolution and social innovation will be positively affected. More concretely I expect:

(i) the introduction of operational definitions for creativity as well as the role of the adjacent possible will allow many different disciplines to see the whole subject in a new light, by providing new analytical tools to make of creativity the object of solid and systematic scientific investigations. In this way new academic audience, also thanks to the large amount of data on creativity and innovation processes the project will make available, could be attracted by the questions raised in this project and the role of creativity could be included in a progressively larger number of course curricula.

(ii) the characterization of the determinants of creativity, as well as the new modeling schemes devised, may very likely trigger a new wave of researches and activities aimed at finding in very different contexts the factors that enhance creativity as well as the best ecologies for them. This in turn could stimulate a research agenda to rethink the spaces we live in and interact, e.g. workplaces, schools, houses, public places, as well as the way in which we organize our interactions and identify best practises for the betterment of our societies.

(iii) as a corollary of the previous point I expect this project could trigger a deep rethinking of educational and learning processes. Through this project I will highlight the red thread linking creativity, games and learning. Learning is at the basis of our ability to understand our environment, adapt to it, survive and evolve as a species. Play is the fundamental context for most human learning across ages: activities like exploration, experimentation, innovation, testing the limits, etc. all happen best when playing. From this perspective I would expect on the long term a major redesign of learning paths to make them more creativity driven.

(iv) I expect this project will stimulate academic and non-academic milieux to investigate the subtle relation between little novelties, those that are part of our everyday lives (we meet new people, hear new words, listen to new songs, use new technologies), and actual innovations. The statistical commonalities we observed so far suggest the hypothesis that novelties and innovations are two sides of the same coin, and I expect this idea will be systematically explored in order to understand how much one can learn about innovations, in biological systems, human society, and technology, by observing the dynamics of mundane novelties.

Let me conclude with a little circular argument. One of the central idea of the present project is that of the adjacent possible, intended as the set of possibilities available to individuals, communities, institutions, organisms, productive processes, etc., at a given point in time during their evolution. From this perspective it is tempting to speculate how the opportunities listed in the present project could perhaps represent the adjacent possible we are all experiencing right now as scientific community. And along the same line of reasoning one could speculate that, once those opportunities will be concretely seized, new adjacent possibilities will be disclosed in a never ending process of triggering events.

ACKNOWLEDGMENTS

The Kreyon project is funded by the John Templeton under contract n. 51663. I wish to warmly thank Vito D.P. Servedio, Steven H. Strogatz and Francesca Tria for the initial steps in the modelling activity of the adjacent possible as well as many enlightening and stimulating discussions. I also wish to thank the whole Kreyon team for its invaluable support both in the definition of the proposal and in its actual implementation: Cesare Bianchi, Indaco Biazzo, Christine Cuskley, Elisabetta Falivene, Pietro Gravino, Anna Lo Piano, Rachele Lo Piano, Bernardo Monechi, Giovanna Chiara Rodi. I'm very proud to thank Ilan Chabay, Andreas Roepstorff, Luc Steels, Bo Stjerne Thomsen who accepted to sit in the Scientific Advisory Board of the project. I'd like to warmly thank all the Institutions that are supporting the project: Sapienza University of Rome, Institute for Scientific Interchange (ISI), the CNR Institute for Complex Systems. Also I'm very grateful to SONY Computer Science Lab in Paris for its kind hospitality during my sabbatical leave in the early stages of this project.

[1] S. Johnson. *Where Good Ideas Come From: The Natural History of Innovation.* Riverhead Hardcover, 2010.
[2] F. Jacob. Evolution and tinkering. *Science*, 196:1161–1166, 1977.
[3] S.J. Gould and E.S. Vrba. Exaptation. A missing term in the science of form. *Annales de l'I.H.P.*, 8(1):4–15, 1982.
[4] Stuart A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford University Press, New York, 1993.
[5] S.M. Reader and K.N. Laland. *Animal Innovation.* Oxford University Press, New York, 2003.
[6] M. Csikszentmihalyi. *Creativity: Flow and the Psychology of Discovery and Invention.* Harper Perennial, New York, 1997.
[7] Everett M. Rogers. *Diffusion of innovations.* Free Press, New York, NY [u.a.], 5th edition, 08 2003.
[8] M. J. O'Brien and S. J. Shennan. *Innovation in Cultural Systems: Contributions from Evolutionary Anthropology.* Vienna Series in Theoretical Biology. MIT Press, 2009.
[9] J. Schumpeter. *The Theory of Economic Development.* Harvard University Press, Cambridge, Mass., 1934.
[10] W.B. Arthur. Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*, 99(394):116–31, March 1989.
[11] G. Dosi, Y. Ermoliev, and Y. Kaniovski. Generalized urn schemes and technological dynamics. *Journal of Mathematical Economics*, 23(1):1–19, 1994.
[12] J.M. Ziman, editor. *Technological Innovation as an Evolutionary Process.* Cambridge University Press, 2000.
[13] W.B. Arthur. *The Nature of Technology.* The Free Press, 2009.
[14] R.V. Solé, S. Valverde, M.R. Casals, S.A. Kauffman, D. Farmer, and N. Eldredge. The evolutionary ecology of technological innovations. *Complexity*, 18(4):15–27, 2013.
[15] Benjamin F. Jones and Bruce A. Weinberg. Age dynamics in scientific creativity. *Proceedings of the National Academy of Sciences*, 108(47):18910–18914, 2011.
[16] H.A. Simon. Discovery, invention, and development: Human creative thinking. *Proc. Natl. Acad. Sci. USA*, 80:4569, 1983.
[17] M.I. Stein. Creativity and culture. *The Journal of Psychology: Interdisciplinary and Applied*, 36(2):311–322, 1953.
[18] F. Barron. The disposition toward originality. *The Journal of Abnormal and Social Psychology*, 51(3):478–485, 1955.
[19] M.A. Runco and G.J. Jaeger. The standard definition of creativity. *Creativity Research Journal*, 24(1):92–96, 2012.
[20] J.C. Houtz and D. Krug. Assessment of creativity: Resolving a mid-life crisis. *Educational Psychology Review*, 7(3):269–300, 1995.
[21] R.J. Sternberg. *Handbook of creativity.* Cambridge Univ. Press, Cambridge, 12. printing. edition, 2009.
[22] E. Villalba. Searching for the holy grail of measuring creativity. *Creativity Research Journal*, 24(1):1–2, 2012.
[23] Dean Keith K. Simonton. Quantifying creativity: can measures span the spectrum? *Dialogues in clinical neuroscience*, 14(1):100–104, 2012.
[24] Denise J. Cai, Sarnoff A. Mednick, Elizabeth M. Harrison, Jennifer C. Kanady, and Sara C. Mednick. Rem, not incubation, improves creativity by priming associa-

tive networks. *Proceedings of the National Academy of Sciences*, 106(25):10130–10134, 2009.

[25] Ruth Ann Atchley, David L. Strayer, and Paul Atchley. Creativity in the wild: Improving creative reasoning through immersion in natural settings. *PLoS ONE*, 7(12):e51474, 12 2012.

[26] Mirta F. Villarreal, Daniel Cerquetti, Silvina Caruso, Violeta Schwarcz López Aranguren, Eliana Roldán Gerschcovich, Ana Lucía Frega, and Ramón C. Leiguarda. Neural correlates of musical creativity: Differences between high and low creative subjects. *PLoS ONE*, 8(9):e75427, 09 2013.

[27] F.M. Bass. A new product growth for model consumer durables. *Manage. Sci.*, 15(1):215–227, 1969.

[28] T.W. Valente. *Network Models of the Diffusion of Innovations.* Quantitative Methods in Communication. Hampton Press, Cresskill, N.J., 1995.

[29] Duncan J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[30] M.J. Salganik, P.S. Dodds, and D.J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856, 2006.

[31] Pierpaolo Andriani and Jack Cohen. From exaptation to radical niche construction in biological and technological complex systems. *Complexity*, 18(5):7–14, 2013.

[32] H. Tettelin, D. Riley, C. Cattuto, and D. Medini. Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, 11(5):472–477, 2008.

[33] A. Robinson (ed.). *Exceptional creativity in science and technology: Individuals, Institutions and Innovations.* Templeton Press, 2013.

[34] K. Robinson. *Out of Our Minds: Learning to be Creative.* Wiley, 2011.

[35] G. Polya. *How to Solve It.* Princeton University Press, November 1971.

[36] David Gauntlett and Bo Stjerne Thomsen. Cultures of creativity: Nurturing creative mindsets across cultures, 2013.

[37] Dean K. Simonton (ed.). *The Wiley Handbook of Genius.* Wiley-Blackwell, 2014.

[38] Stuart A. Kauffman. *Investigations.* Oxford University Press, New York/Oxford, 2000.

[39] F. Jacob. *Le jeu des possibles.* Fayard, 1981.

[40] S.A. Kauffman, S. Thurner, and R. Hanel. The evolving web of future wealth. *Scientific American (online)*, 2008.

[41] S. Thurner, P. Klimek, and R. Hanel. Schumpeterian economic dynamics as a quantifiable minimum model of evolution. *New Journal of Physics*, 12:075029, 2010.

[42] Teppo Felin, Stuart A. Kauffman, Roger Koppl, and Giuseppe Longo. Economic Opportunity and Evolution: Beyond Bounded Rationality and Phase Space. *Strategic Entrepreneurship Journal*, 8(4):269–282, 2014.

[43] M. Buchanan. Great leap outwards. *Nature Physics*, 10:243, 2014.

[44] Francesca Tria, Vittorio Loreto, Vito Domenico Pietro Servedio, and Steven H. Strogatz. The dynamics of correlated novelties. *Nature Scientific Reports*, 4, 2014.

[45] G. Pólya. Sur quelques points de la théorie des probabilités. *Annales de l'I.H.P.*, 1(2):117–161, 1930.

[46] N. L. Johnson and S. Kotz. *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory.* Wiley, 1977.

[47] H. Mahmoud. *Pólya Urn Models.* Texts in Statistical Science series. Taylor and Francis Ltd, Hoboken, NJ, 2008.

[48] Udny G. Yule. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character:*, 213:21–87, 1925.

[49] H.A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3-4):425–440, 1955.

[50] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1:226–251, 2003.

[51] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46:323–351, December 2005.

[52] M.V. Simkin and V.P. Roychowdhury. Re-inventing Willis. *Physics Reports*, 502(1):1 – 35, 2011.

[53] A. Robinson. *Sudden Genius?: The Gradual Path to Creative Breakthroughs.* OUP Oxford, 2010.

[54] D. Benedetto, E. Caglioti, and V. Loreto. Language trees and zipping. *Phys. Rev. Lett.*, 88:048702, Jan 2002.

[55] M. Li and P.M.B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer Publishing Company, Incorporated, 3 edition, 2008.

[56] J.B. Michel, Y.-K. Shen, A.P. Aiden, A. Veres, M.K. Gray, The Google Books Team, J.P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M.A. Nowak, and E. Aiden Lieberman. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182, 2011.

[57] T. Bertin-Mahieux, D.P.W. Ellis, B. Whitman, and P. Lamere. The million song dataset. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, 2011.

[58] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326, New York, NY, USA, 2004. ACM Press.

[59] Luis von Ahn. Games with a purpose. *Computer*, 39(6):92–94, 2006.

[60] Matthew J. Salganik and Duncan J. Watts. Web-Based Experiments for the Study of Collective Social Dynamics in Cultural Markets. *Topics in Cognitive Science*, 1(3):439–468, 2009.

[61] S. Caminiti, C. Cicali, P. Gravino, V. Loreto, V.D.P. Servedio, A. Sirbu, and F. Tria. Xtribe: A web-based social computation platform. In *Cloud and Green Computing (CGC), 2013 Third International Conference on*, pages 397–403, Sept 2013.

[62] A. Baronchelli, V. Loreto, and F. Tria, editors. *Language Dynamics*, volume 15 of *Special Issue of Advances in Complex Systems.* World Scientific, 2012.

[63] L. Wittgenstein. *Philosophical Investigations. (Translated by Anscombe, G.E.M.).* Basil Blackwell, Oxford, UK, 1953.