ISSN 2427-4577

BULLETIN N° 256
ACADÉMIE EUROPÉENNE INTERDISCIPLINAIRE DES SCIENCES
INTERDISCIPLINARY EUROPEAN ACADEMY OF SCIENCES

Lundi 4 avril 2022 (en format mixte présence-distance) :

15h30 : Conférence

« Crises économiques : un point de vue de physicien »

Par **Jean-Philippe BOUCHAUD**
Professeur à l'École Normale Supérieure
Membre de l'Académie des Sciences
Président et directeur de la recherche de Capital Fund Management (CFM), Paris

Notre Prochaine séance aura lieu le lundi 2 mai 2022 de 15h30 à 18h00
Salle Annexe Amphi Burg
Institut Curie, 13 rue Lhomond - 75005 Paris

Elle aura pour thème :

15h30 : Conférence

« ***Vers une médecine computationnelle de précision*** »

Par **Mickaël GUEDJ**
Directeur de la biométrique,
des sciences des données et de la décision
chez Nanobiotix

ACADÉMIE EUROPÉENNE INTERDISCIPLINAIRE DES SCIENCES
INTERDISCIPLINARY EUROPEAN ACADEMY OF SCIENCES

Avril 2022

N°256

TABLE DES MATIERES

Prochaine séance : lundi 2 mai 2022 de 15h30 à 18h00

15h30 : Conférence

« ***Vers une médecine computationnelle de précision*** »

Par **Mickaël GUEDJ**
Directeur de la biométrique,
des sciences des données et de la décision
chez Nanobiotix

ACADÉMIE EUROPÉENNE INTERDISCIPLINAIRE DES SCIENCES

INTERDISCIPLINARY EUROPEAN ACADEMY OF SCIENCES

Séance du Lundi 4 avril 2022 mixte présentiel-distanciel

La séance est ouverte à 15h30, **sous la Présidence de Victor MASTRANGELO**

- **Étaient présents physiquement nos Collègues membres titulaires** de Paris : Gilbert BELAUBRE, Jean BERBINAU, Éric CHENIN, Françoise DUTHEIL, Irène HERPE-LITWIN, Jean SCHMETS, Jean-Pierre TREUIL, ainsi que notre Président de la section d'Athènes, Anastassios METAXAS
- **Etaient présent physiquement notre Collègue membre correspondant** Benoit PRIEUR
- **Étaient connectés à distance nos Collègues** : François BOUCHET, Ernesto DI MAURO, Michel GONDRAN, Christian GORINI, Dominique PRAPOTNITCH, Édith PERRIER, Jacques PRINTZ,
- **Était excusée notre Collègue :** Anne BURBAN (en mission à Beyrouth)

I. Conférence du Pr Jean-Philippe BOUCHAUD : *Crises économiques : un point de vue de Physicien*

1. Présentation du Conférencier par notre Président Victor MASTRANGELO

Jean-Philippe BOUCHAUD est un ancien élève de l'École Normale Supérieure de la rue d'Ulm. Il a préparé son doctorat de Physique au Laboratoire de Spectroscopie Hertzienne de l'ENS et l'a soutenue en 1985.Il est spécialiste de physique statistique des milieux désordonnés et l'un des pionniers de l'« ***éconophysique***», discipline qui cherche à appliquer les concepts et les méthodes de la physique aux systèmes économiques et aux marchés.

Activités actuelles :

Depuis l'an 2000, il est Président et directeur de la recherche de CFM : **Capital Fund Management (CFM)** et responsable avec Marc POTTERS d'un groupe de recherche de plusieurs dizaines de PhDs.

- Depuis 2009, il est *Président de la Fondation CFM pour la Recherche*
- Depuis 2017, il est *Professeur au Master 2 de physique de l'ENS*, en charge du cours : « *D*e *la physique statistique aux sciences sociales* » - « *From Statistical Physics to Social Sciences* »
- Depuis 2014, il est Co-Head du "*CFM-Imperial Institute for quantitative finance*"
- Il est éditeur de "*Market Microstructure & Liquidity* »

Activités passées :

- De 1985-1992 il a été *Chargé de recherche au CNRS*
- De 1993-96 il a été *Ingénieur de recherché au SPEC, CEA-Saclay*
- De 1996-2006 il a été *Expert Sénior au CEA-Saclay*
- Il a été en 2001, *Chercheur invité à Harvard et en charge de deux cours (Harvard & MIT)*
- De 2003-2012 il a été *Coéditeur en chef avec Michael Dempster de Quantitative Finance*
- De 2009-2017 il a été *Professeur de Physique Statistique & « Systèmes Complexes » à l'Ecole Polytechnique*

Prix et honneurs :

- En 1996, le *CNRS lui a attribué la Médaille d'Argent* pour ses travaux
- le Risk Magazine **l'a désigné comme** *Quant of the Year* pour les années 2017 et 2018
- En 2018, il a été élu à l'Académie des Sciences

Il a encadré ou co-encadré (depuis 1992) une vingtaine de thèses de doctorat :

Publications

Livres:
* Théorie des Risques Financiers, with M. Potters, Eyrolles (Paris) 1997
* Theory of Financial Risks and Derivative Pricing, with M. Potters, Cambridge University Press (2000 and 2003)
* Lévy flights and Laser cooling, with A. Aspect, F. Bardou, C. Cohen-Tannoudji, Cambridge University Press (2001)
* Complex Systems, Les Houches Lecture Notes (2006), Editor with M. Mézard, J. Dalibard, Elsevier (2008)
* Dynamical heterogeneities in glasses, colloids, and granular media, Editor with L. Berthier, G. Biroli, L. Cipelletti, W. van Saarlos, Oxford University Press (2001)
* Market Microstructure, confronting many viewpoints, Editor with Frédéric Abergel, Thierry Foucault, Charles-Albert Lehalle, Wiley 2012
* Trades, Quotes and Prices. Financial Markets Under the Microscope, with J. Bonart, J. Donier and M. Gould, Cambridge University Press (2018)

Depuis 1984 il a à son actif 360 publications scientifiques dans des revues internationales à comité de lecture dans différents domaines (Physique Statistique, Verres et Verres de Spin, Matière Granulaire, Théorie des Matrices Aléatoires, Finance Quantitative, Macroéconomie). H-index (Google Scholar):

Il a également à son actif une dizaine d'articles dans des revues internationales ainsi que des chapitres de livre**.**

Brevet: Evaluating and optimizing error-correcting codes using a renormalization group transformation JS Yedidia, JP Bouchaud US Patent 6,857,097 857,097

En décembre 2016, L'Ecole normale supérieure a inauguré une nouvelle Chaire « Modèles et Sciences des Données » grâce à un partenariat avec la Fondation CFM pour la Recherche.

En décembre 2019 une chaire de recherche « Econophysique & Systèmes Complexes » a été créée à l'Ecole polytechnique grâce à un partenariat **avec la Fondation CFM pour la Recherche** et la Fondation du Risque (FDR) de l'Institut Louis Bachelier

Enfin en 2020-2021 il a été invité sur la chaire annuelle ***Innovation technologique – Liliane Bettencourt*** du Collège de France.

Le jeudi 25 février 2021 il a prononcé sa leçon inaugurale sur le thème **« *De la physique statistique aux sciences sociales : les défis de la pluridisciplinarité* ».**

2. Résumé de la conférence

« Crises Economiques : Un point de vue de physicien »

Résumé

Comme l'écrivait P.W. Anderson en 1972 dans son article « *More is different* », le comportement de grandes assemblées d'individus (ou de molécules) ne peut pas être compris à partir de l'extrapolation du comportement d'individus (ou de molécules) isolé(e)s. Au contraire, des comportements complètement nouveaux, parfois spectaculaires et difficiles à anticiper, peuvent apparaitre et nécessitent des idées et des méthodes nouvelles. L'objet de la physique statistique est précisément de tenter de comprendre ces phénomènes collectifs, qui n'appartiennent à aucun des constituants élémentaires sous-jacents.
En particulier, de petits changements au niveau individuel peuvent entrainer des effets dramatiques au niveau collectif. Plusieurs exemples simples seront discutés, qui démontrent la nécessité d'aller au-delà des modèles de l'économie classique, basés sur l'idée d'un « agent représentatif » (de surcroît rationnel), et pour lesquels ***seuls des évènements exogènes peuvent conduire à des crises – alors que de nombreux phénomènes socio-économiques ou financiers semblent de nature endogène.***

Abstract

As P.W. Anderson wrote in 1972 in his paper « *More is different* », the behaviour of large gatherings of individuals (or molecules) cannot be explained through the extrapolation of the behaviour of separate individuals (or molecules). On the contrary, completely novel behaviours, sometimes spectacular and difficult to anticipate, can appear and require new ideas and methods. The subject of Statistical Physics is precisely to try and understand these collective phenomena, which do not belong to any of the underlying elementary components.

In particular, slight changes at the individual level may have dramatic impacts at the collective level. Several simple cases will be discussed which show the necessity to go beyond the models of classical economy based on the idea of a « representative agent » (in addition, rational), and for whom, only exogenous events can lead to crises – although numerous socio-economic or financial phenomena seem to have an endogenous nature.

3. Exposé du conférencier et échanges avec l'auditoire :

Une synthèse complète de la conférence et des échanges est en cours d'élaboration par notre Collègue Jean BERBINAU. Elle sera très prochainement accessible dans la rubrique Comptes-rendus conférences mensuelles du site de l'AEIS http://www.science-inter.com

Néanmoins, comme nos Collègues Jean Pierre TREUIL, Jean BERBINAU, Benoist PRIEUR, Irène HERPE-LITWIN, Edith PERRIER, Ernesto DI MAURO et notre Président Victor MASTRANGELO ont posé de nombreuses questions., nous vous proposons ci-dessous une esquisse de ces échanges. Ceux-ci seront traités plus en détail dans la synthèse complète de la conférence et des échanges mentionnée ci-dessus.

Résumé de l'échange qui a suivi la conférence :

Jean-Pierre TREUIL questionne la capacité explicative des modèles multi-agents : même lorsqu'ils reproduisent des comportements collectifs observés dans la réalité, ils ne fournissent pas de formalisme explicatif satisfaisant. Jean-Philippe BOUCHAUD abonde dans son sens : il considère que les modèles multi-agents permettent de faire des expériences in silico et d'observer les comportements, comme dans la réalité, pourvu que le modèle soit juste ; reste ensuite à essayer de comprendre ce que l'on a observé : à identifier les formalismes explicatifs sous-jacents, par exemple des boucles de rétroaction.

Victor MASTRANGELO aborde la non-linéarité des phénomènes simulés, et les précautions qu'il faut prendre de ce fait en matière de critères de convergence pour s'assurer de coller au mieux à la réalité. Jean-Philippe BOUCHAUD indique qu'un modèle approximativement vrai permet quand-même d'étudier les mécanismes impliqués.

Jean BERBINAU aborde la difficulté de prédire lorsque les phénomènes sont trop aléatoires et les domaines de variation potentielle trop importants, ainsi que son corollaire : la difficulté de convaincre de la pertinence des prédictions lorsque celles-ci s'écartent trop des comportements courants (exemple de la crise de 2008). Jean-Philippe BOUCHAUD souligne qu'on prédit plutôt des distributions de probabilités et des listes d'événements possibles, que des valeurs précises des variables observées.

Jean-Pierre TREUIL se demande ensuite si le comportement collectif d'agents conscients peut être appréhendé avec des lois physiques. Jean-Philippe BOUCHAUD indique que très peu des caractéristiques du comportement individuel des agents transparaissent au niveau collectif. Pour des idées ou des opinions par exemple, on retrouve au niveau collectif des phénomènes de propagation analogues à ce qui se passe en physique (diffusion) ou en biologie (virus).

Irène HERPE-LITWIN trouve intéressants ces phénomènes de massification de l'opinion où chacun ne raisonne plus par lui-même mais suit un courant commun (par exemple, la doxa officielle), ce qui ouvre vers un régime totalitaire.

Benoît PRIEUR évoque la théorie du chaos et la difficulté intrinsèque de modéliser un système économique réel, qui intègre de nombreux biais, des comportements malhonnêtes, des prédictions autoréalisatrices. Comment peut-on modéliser des systèmes qui intègrent autant de paramètres hétérogènes ? Jean-Philippe BOUCHAUD indique que chaotique ne signifie pas totalement imprévisible : le système peut être imprévisible au sens déterministe, mais il a des propriétés statistiques qui restent prévisibles. En revanche, des systèmes ayant la complexité du réel en matière économique sont pires : leurs probabilités elles-mêmes sont non connaissables. Pour de tels systèmes, la liste des scénarios possibles est déjà un but en soi.

Jean-Pierre TREUIL se demande si la capacité d'anticipation des agents conscients change le comportement global. Jean-Philippe BOUCHAUD indique qu'il est difficile d'anticiper les décisions des autres agents ou les émergences (par exemple l'évènement du téléphone portable) dans la vie réelle, et que l'anticipation joue de ce fait un rôle limité.

Edith PERRIER trouve que cela fait longtemps maintenant que l'on a découvert la modélisation à base d'agents, que l'on sait simuler des comportements collectifs et que l'on a mis en évidence la notion d'émergence. Jean-Philippe BOUCHAUD estime que rien n'a encore été fait pour appliquer ces approches de façon massive en économie. Même en météorologie, les méthodes numériques ont eu du mal à percer. Et l'économie est un des domaines les plus isolés scientifiquement : avec le moins de citations croisées entre l'économie et les autres domaines. Un élément prometteur, pour J. Ph. BOUCHAUD, est le développement actuel de la science des données et l'apparition des données massives dans toutes les activités ; encore faut-il que les économistes saisissent cette opportunité.

Jacques PRINTZ s'intéresse au domaine particulier des projets, qui constituent des communautés de quelques centaines à quelques dizaines de milliers de personnes. Il évoque le rapport du « *standish group* », intitulé chaos, qui concluait que 2/3 des projets échouaient. L'analyse indiquait que les échecs étaient dus à de petits

événements passés inaperçus tôt dans le projet, et que lorsque la perspective de l'échec s'annonçait, les mesures prises comprenaient généralement un renfort en personnel, lequel souffrait souvent d'un déficit de qualification et de connaissance du projet, et de ce fait sur-mobilisait les personnes compétentes, et ce qui freinait encore plus l'avancement du projet au lieu de l'accélérer, et aggravait les difficultés au lieu de les résoudre. Jacques PRINTZ se demande si les approches de modélisation exposées peuvent aider dans ces cas-là. Jean-Philippe BOUCHAUD indique que la notion de boucle de rétroaction, élément clé des approches exposées, peut être utile pour anticiper les difficultés et identifier les solutions dans certains cas.

Ernesto DI MAURO évoque l'influence de l'échelle d'un système : a-t-elle une influence qualitative sur le fonctionnement du système ? Jean-Philippe BOUCHAUD estime que les phénomènes de grande échelle apparaissent généralement assez vite quand la taille croît : on atteint assez vite une taille suffisante pour avoir un comportement analogue à celui d'un système de très grande taille. Par exemple, pour des spins en interaction, la magnétisation, avec suffisamment de temps, se produit comme dans un système de très grande taille, dès qu'on a quelques centaines de spins. Généralement, le passage à la limite se fait rapidement. Donc les réponses qualitatives aux questions que l'on se pose ne nécessitent pas un très grand nombre d'agents. Ernesto poursuit en se demandant quelles différences qualitatives il y a entre les systèmes économiques de pays de tailles très différentes (exemple : Lichtenstein et Etats-Unis). Jean-Philippe BOUCHAUD indique qu'une petite taille peut induire une structure très particulière avec un tout petit nombre d'acteurs économiques qui ont une influence prépondérante sur le comportement global (exemple de Nokia en Finlande).

Enfin, Victor MASTRANGELO se demande si dans le domaine économique, on peut identifier des paramètres qui jouent un rôle analogue à celui de la température en physique. Jean-Philippe BOUCHAUD indique que dans les choix des agents, intervient une part de rationalité, et une part d'aléatoire. Cet aléatoire joue un rôle analogue à celui de la température. Plus d'aléatoire conduit à une configuration homogène, analogue à la phase gazeuse ; moins d'aléatoire conduit à une configuration plus hétérogène, analogue à une phase solide (cristallisation).

Sont également disponibles dans la rubrique Comptes-rendus des conférences mensuelles du site de l'AEIS http://www.science-inter.com les enregistrements en visio-conférences et les documents suivants pour la conférence de Jean-Philippe BOUCHAUD :

- Conférence (à venir)
- Présentation du conférencier (à venir)
- Résumé (.pdf)
- Support de présentation (.pdf)
- Synthèse de la conférence et des échanges qui ont suivi (.pdf) (à venir)
- Article associé 1 (.pdf) : « *De la physique pour analyser les crises économiques* », de Jean-Philippe Bouchaud, *Collection Idées de l'Académie des Sciences.* A partir d'outils scientifiques issus de la physique des systèmes complexes, Jean-Philippe Bouchaud montre en quoi les théories classiques s'avèrent inadéquates pour comprendre et prévenir les krachs.
- Article associé 2 (.pdf) : « *More Is Different* », de P. W. Anderson, *Revue Science, volume 177, numéro 4047, du 4 Août 1972.* Broken symmetry and the nature of the hierarchical structure of science.
- Article associé 3 (.pdf) : « *Tipping points in macroeconomic agent-based models* », de Gualdi, S., Tarzia, M., Zamponi, F., & Bouchaud, J., *Journal of Economic Dynamics and Control*, *50*, 29-61
- Article associé 4 (.pdf) : « *Optimal inflation target : insights from an Agent-Based Model* », de Jean-Philippe Bouchaud et al., *intervention au congrès NAEC, 21 Janvier 2020.*

REMERCIEMENTS
Nous tenons à remercier vivement M. Yann TRAN et Mme Annabelle POIRIER de l'Institut Curie pour la qualité de leur accueil.

II. Conférence de Mickaël GUEDJ :

Vers une Médecine Computationnelle de Précision

Mickaël GUEDJ, PhD
Head of Biometrics, Data & Decision Sciences, Nanobiotix

Résumé

Vers une Médecine Computationnelle de Précision
Les Sciences Computationnelles (incluant l'Intelligence Artificielle) s'appuient sur une convergence de technologies offrant des synergies avec les technologies des Sciences de la Vie afin de capturer la valeur de données biomédicales massives sous la forme de modèles prédictifs et soutenant la prise de décision. Les algorithmes optimisent ainsi la découverte et le développement de médicaments en améliorant notre compréhension de l'hétérogénéité des maladies, en identifiant les voies moléculaires dérégulées, les cibles thérapeutiques et les médicaments candidats. Ce niveau de connaissances sans précédent concernant les spécificités des patients favorise l'émergence d'une Médecine Computationnelle de Précision permettant la conception de thérapies adaptées aux singularités de chaque patient en termes de physiologie et de caractéristiques de la maladie.

Abstract

Toward a Computational Precision Medicine
Computational Sciences including Artificial Intelligence (AI) relies upon a convergence of technologies with further synergies with Life sciences technologies to capture the value of big biomedical data in the form of predictive models supporting decision-making. Algorithms enhance drug discovery and development by improving our understanding of disease heterogeneity, identifying dysregulated molecular pathways, therapeutic targets and drug candidates. This unprecedented level of knowledge on patient specificities is fostering the emergence of a Computational Precision Medicine allowing the design of therapies tailored to the singularities of individual patients in terms of their physiology and disease features.
Mickael Guedj has 20 years experience in the treatment of large-scale biomedical data to support drug discovery & development. With a special interest for understanding pathophysiological mechanisms, new therapeutic target identification and drug repurposing.

Publications

1) **Artificial intelligence-enhanced drug design and development: Toward a computational precision medicine** paru chez Drug Discovery Today par Philippe MOINGEON et al. accessible sur le site https://www.sciencedirect.com/science/article/abs/pii/S1359644621003962
2) **A new molecular classification to drive precision treatment strategies in primary Sjögren's syndrome**, paru chez Nature Communications par Perrine SORET et al. accessible sur le site https://doi.org/10.1038/s41467-021-23472-7
3) **Network-based repurposing identifies anti-alarmins as drug candidates to control severe lung inflammation in COVID-19** par Emiko DESVAUX et al. paru chez PLOS ONE accessible sur le site https://doi.org/10.1371/journal.pone.0254374

Eléments de CV de **Mickaël GUEDJ**

Mickael Guedj est diplomé de l'INSA-Lyon en Bioinformatique & Modélisation et a une thèse en Statistique Génétique réalisée au sein de la Génopole d'Evry. Il a passé ensuite 2 ans à la Ligue Nationale contre le Cancer, 8 ans chez Pharnext, 4 ans chez Servier et depuis 6 mois chez Nanobiotix, spécialisé dans le traitement des biomédicales à grande échelle pour soutenir la découverte et le développement de médicaments. Il a un intérêt particulier pour la compréhension des mécanismes physiopathologiques, l'identification de nouvelles cibles thérapeutiques et le repositionnement de médicaments.

Documents

Pour compléter la conférence du Pr Jean-Philippe BOUCHAUD nous vous proposons :

p. 11 Un article de 2015 intitulé « *Tipping points in macroeconomic agent-based models* » de Gualdi, S., Tarzia, M., Zamponi, F., & Bouchaud, J. paru dans *Journal of Economic Dynamics and Control*, *50*, 29-61

Pour préparer la conférence de Mickaël GUEDJ, nous vous proposons :

p. 53 un article sous presse chez Elsevier intitulé « *Artificial intelligence-enhanced drug design and development: Toward a computational precision medicine"* de Philippe Moingeon, Melaine Kuenemann, et Mickaël Guedj

p.61 Un article intitulé « *Network-based repurposing identifies anti-alarmins as drug candidates to control severe lung inflammation in COVID-19* » par Emiko DESVAUX et al. paru chez PLOS ONE accessible sur le site https://doi.org/10.1371/journal.pone.0254374

Tipping points in macroeconomic Agent-Based models

Stanislao Gualdi,[1] Marco Tarzia,[2] Francesco Zamponi,[3] and Jean-Philippe Bouchaud[4]

[1] *Université Pierre et Marie Curie - Paris 6, Laboratoire de Physique Théorique de la Matière Condensée, 4, Place Jussieu, Tour 12, 75252 Paris Cedex 05, France and IRAMIS, CEA-Saclay, 91191 Gif sur Yvette Cedex, France* *

[2] *Université Pierre et Marie Curie - Paris 6, Laboratoire de Physique Théorique de la Matière Condensée, 4, Place Jussieu, Tour 12, 75252 Paris Cedex 05, France*

[3] *Laboratoire de Physique Théorique, École Normale Supérieure, UMR 8549 CNRS, 24 Rue Lhomond, 75231 Paris Cedex 05, France*

[4] *CFM, 23 rue de l'Université, 75007 Paris, France, and Ecole Polytechnique, 91120 Palaiseau, France.*

The aim of this work is to explore the possible types of *phenomena* that simple macroeconomic Agent-Based models (ABM) can reproduce. We propose a methodology, inspired by statistical physics, that characterizes a model through its "phase diagram" in the space of parameters. Our first motivation is to understand the large macro-economic fluctuations observed in the "Mark I" ABM devised by D. Delli Gatti and collaborators. In this regard, our major finding is the generic existence of a *phase transition* between a "good economy" where unemployment is low, and a "bad economy" where unemployment is high. We then introduce a simpler framework that allows us to show that this transition is *robust* against many modifications of the model, and is generically induced by an asymmetry between the rate of hiring and the rate of firing of the firms. The unemployment level remains small until a tipping point, beyond which the economy suddenly collapses. If the parameters are such that the system is close to this transition, any small fluctuation is amplified as the system jumps between the two equilibria. We have explored several natural extensions of the model. One is to introduce a bankruptcy threshold, limiting the firms maximum level of debt-to-sales ratio. This leads to a rich phase diagram with, in particular, a region where *acute endogenous crises* occur, during which the unemployment rate shoots up before the economy can recover. We also introduce simple wage policies. This leads to inflation (in the "good" phase) or deflation (in the "bad" phase), but leaves the overall phase diagram of the model essentially unchanged. We have also explored the effect of simple monetary policies that attempt to contain rising unemployment and defang crises. We end the paper with general comments on the usefulness of ABMs to model macroeconomic phenomena, in particular in view of the time needed to reach a steady state that raises the issue of *ergodicity* in these models.

It is human nature to think wisely and to act absurdly – Anatole France

*Electronic address: stanislao.gualdi@gmail.com

Contents

I. INTRODUCTION

A. From micro-rules to macro-behaviour

Inferring the behaviour of large assemblies from the behaviour of its elementary constituents is arguably one of the most important problems in a host of different disciplines: physics, material sciences, biology, computer sciences, sociology and, of course, economics and finance. It is also a notoriously hard problem. Statistical physics has developed in the last 150 years essentially to understand this micro-macro link. Clearly, when interactions are absent or small enough, the system as a whole merely reflects the properties of individual entities. This is the canvas of traditional macro-economic approaches. Economic agents are assumed to be identical, non-interacting, rational agents, an idealization known as the "Representative Agent" (RA). In this framework, micro and macro trivially coincide. However, we know (in particular from physics) that discreteness, heterogeneities and/or interactions can lead to totally unexpected phenomena. Think for example of super-conductivity or super-fluidity[1]: before their experimental discovery, it was simply beyond human imagination that individual electrons or atoms could "conspire" to create a collective state that can flow without friction. Micro and macro behaviour not only do not coincide in general, but genuinely *surprising* behaviour can emerge through aggregation. From the point of view of economic theory, this is interesting, because financial and economic history is strewn with bubbles, crashes, crises and upheavals of all sorts. These are very hard to fathom within a Representative Agent framework [2], within which crises would require large aggregate shocks, when in fact small local shocks can trigger large systemic effects when heterogeneities, interactions and network effects are taken into account [3–6]. The need to include these effects has spurred a large activity in "Agent-Based models" (ABMs) [7–9]. These models need numerical simulations, but are extremely versatile because any possible behavioural rules, interactions, heterogeneities can be taken into account.

In fact, these models are so versatile that they suffer from the "wilderness of high dimensional spaces" (paraphrasing Sims [10]). The number of parameters and explicit or implicit choices of behavioural rules is so large (~ 10 or more, even in the simplest models, see below) that the results of the model may appear unreliable and arbitrary, and the calibration of the parameters is an hopeless (or highly unstable) task. Mainstream RA "Dynamic Stochastic General Equilibrium" models (DSGE), on the other hand, are simple enough to lead to closed form analytical results, with simple narratives and well-trodden calibration avenues [11]. In spite of their unrealistic character, these models appear to perform satisfactorily in 'normal' times, when fluctuations are small. However, they become deeply flawed in times of economic instability [12], suggesting different assumptions are needed to understand what is observed in reality. But even after the 2008 crisis, these traditional models are still favoured by most economists, both in academia and in institutional and professional circles. ABMs are seen at best as a promising research direction and at worst as an unwarranted "black box" (see [13] for an enlightening discussion on the debate between traditional DSGE models and ABMs, and [14–17] for further insights).

B. A methodological manifesto

At this stage, it seems to us that some clarifications are indeed needed, concerning both the objectives and methodology of Agent-Based models. ABMs do indeed suffer from the wilderness of high dimensional spaces, and some guidance is necessary to put these models on a firm footing. In this respect, statistical physics offers a key concept: the *phase diagram* in parameter space [18]. A classic example, shown in Fig. 1, is the phase diagram of usual substances as a function of two parameters, here temperature and pressure. The generic picture is that the number of distinct phases is usually small (e.g. three in the example of Fig. 1: solid, liquid, gas). Well within each phase, the properties are qualitatively similar and small changes of parameters have a small effect. Macroscopic (aggregate) properties do not fluctuate any more for very large systems and are robust against changes of microscopic details. This is the "nice" scenario, where the dynamics of the system can be described in terms of a small number of macroscopic (aggregate) variables, with some effective parameters that encode the microscopic details. But other scenarios are of course possible; for example, if one sits close to the boundary between two phases, fluctuations can remain large even for large systems and small changes of parameters can radically change the macroscopic behaviour of the system. There may be mechanisms naturally driving the system close to criticality (like in Self Organized Criticality [19]), or, alternatively, situations in which whole phases are critical, like for "spin-glasses" [20].

In any case, the very first step in exploring the properties of an Agent-Based model should be, we believe, to identify the different possible phases in parameter space and the location of the phase boundaries. In order to do

[1] See e.g. Ref. [1] for an history of the discovery of super-fluidity and a list of references.

this, numerical simulations turn out to be very helpful [21, 22]: aggregate behaviour usually quickly sets in, even for small sizes. Some parameters usually turn out to be crucial, while others are found to play little role. This is useful to establish a qualitative *phenomenology* of the model – what kind of behaviour can the model reproduce, which basic mechanisms are important, which effects are potentially missing? This first, qualitative step allows one to unveil the "skeleton" of the ABM. Simplified models that retain most of the phenomenology can then be constructed and perhaps solved analytically, enhancing the understanding of the important mechanisms, and providing some narrative to make sense of the observed effects. In our opinion, calibration of an ABM using real data can only start to make sense after this initial qualitative investigation is in full command – which is in itself not trivial when the number of parameters is large. The phase diagram of the model allows one to restrict the region of parameters that lead to "reasonable" outcomes (see for example the discussion in [23, 24]).

C. Outline, results and limitations of the paper

The aim of this paper is to put these ideas into practice in the context of a well-studied macroeconomic Agent-Based model (called "Mark I" below), devised by Delli Gatti and collaborators [25, 26]. This model is at the core of the European project "CRISIS", which partly justifies our attempt to shed some theoretical light on this framework.[2] In the first part of the paper, we briefly recall the main ingredients of the model and show that as one increases the baseline interest rate, there is a phase transition between a "good" state of the economy, where unemployment is low and a "bad" state of the economy where production and demand collapse. In the second part of the paper, we study the phase diagram of a highly simplified version of Mark I, dubbed 'Mark 0' that aims at capturing the main drivers of this phase transition. The model is a "hybrid" macro/ABM model where firms are treated individually but households are only described in aggregate. Mark 0 does not include any exogeneous shock; crises can only be of endogeneous origin. We show that the most important parameter in this regard is the asymmetry between the firms' propensity to hire (when business is good) or to fire (when business is bad). In Mark I, this asymmetry is induced by the reaction of firms to the level of interest rates, but other plausible mechanisms would lead to the same effect. The simplest version of the model is amenable to an analytic treatment and exhibits a "tipping point" (i.e. a discontinuous transition) between high employment and high unemployment. When a bankruptcy condition is introduced (in the form of a maximum level of debt-to-sales ratio), the model reveals an extremely rich phenomenology: the "good" phase of the economy is further split into three distinct phases: one of full employment, a second one with a substantial level of residual unemployment, and a third, highly interesting region, where endogeneous crises appear. We find that both the amount of credit available to firms (which in our model sets the bankrupcty threshold), and the way default costs are absorbed by the system, are the most important aspects in shaping the qualitative behavior of the economy. Finally, we allow wages to adapt (whereas they are kept fixed in Mark I) and allow inflation or deflation to set in. Still, the overall shape of the phase diagram is not modified. We investigate further enhancements of the models, in particular simple policy experiments. Open questions and future directions, in particular concerning macroeconomic ABMs in general, are discussed in our final section.

Before embarking to the core of our results, we want to clearly state what our ambition and objectives are, and what they are not. We do claim that the methodology proposed here is interesting and general, and could help improving the relevance of macroeconomic ABMs. We do believe that large aggregate volatility and crises (in particular those appearing in Mark I) can be understood through the lens of instabilities and phase transitions, as exemplified by our highly stylized Mark 0 model. We are also convinced that simple "skeleton" ABMs (for which most of the phenomenology can be fully dissected) must be developed and compared with traditional DSGE models before embarking into full-fledged models of the economy. On the other hand, we do *not* claim that our basic Mark 0 framework is necessarily the best starting point. Mark 0 was primarily set up as a simplification of Mark I. Still, we find that Mark 0 leads to a surprisingly rich and to some extent realistic set of possible behaviours, including business cycles and crises, inflation, policy experiments, etc. However, we do *not* wish to claim that Mark 0 is able to reproduce all the known empirical stylized facts and could well be in contradiction with some of them[3]. In view of the simplicity of the model, this has to be expected. But we believe that Mark 0 can serve as a useful building block in the quest of a more comprehensive ABM, as more and more effects are progressively incorporated, in a controlled

---

[2] see www.crisis-economics.eu

[3] The idea of building *mathematical models of reality* that reproduce some phenomena, but might even be in contradiction with others, is at the heart of the development of physics. Most probably, its use was introduced in the Hellenistic period [27]. A striking example [27] is Archimedes' *On Floating Bodies*. In the first of the two books, in fact, Archimedes provides a mathematical proof of the sphericity of Earth (assumed to be liquid and at rest). However, in the second book, he assumes the surface of Earth to be flat for the purpose of describing other phenomena, for which the sphericity of Earth is irrelevant.

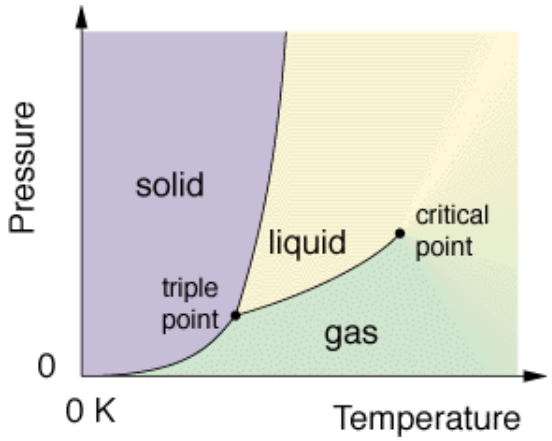

FIG. 1: A typical "phase diagram", here the solid-liquid-gas phases in the temperature-pressure plane. Far from phase transitions, within a given phase, the behaviour of the system is qualitatively similar for all values of the parameters. Close to phase boundaries, on the other hand, small changes of parameters can lead to dramatic macroscopic changes.

manner, to the model. Even if Mark 0 turns out to be little more than a methodological exercice, we hope that the ABM community (and perhaps beyond) will find it inspiring.

To conclude this (long) introduction, let us insist that all the claims made in this paper only refer to the studied *models*, but do not necessarily apply to economic reality. If fact, our central point is that a model has to be understood inside-out before even trying to match any empirical fact.

II. A PHASE TRANSITION IN "MARK I"

A. Description of the model in a nutshell

The Mark I family of agent-based models was proposed by Delli Gatti and collaborators as a family of simple stylized macroeconomic models [25, 26]. Note that several other macroeconomic Agent-Based models have been put forth in the recent years, see [28–33]. Mark I is particularly interesting because large fluctuations in unemployment and output seem to persist in the stationary state (a feature in fact shared by many ABMs cited above).

The Mark I economy [25] is made up of a set of firms, households, firms owners and a bank. Firms produce a certain quantity of a single (and not storable) good, proportional to the number of their employees, that is sold at a time-dependent and firm-dependent prices. Firms pay identical time-independent wages. When the cash owned by a firm is not enough to pay the wages, it asks banks for a loan. The bank provides loans to the firms at an interest rate that depends on the financial fragility of the firm. Households provide workforce in exchange of a salary and want to spend a fixed fraction of their savings (or wealth). The owners of the firms do not work but receive dividends if the firms make profits. Firms are adaptive, in the sense that they continuously update their production (i.e. they hire/fire workers) and their prices, in an attempt to match their production with the demand of goods issued by the households. They also choose how much extra loan they want to take on, as a function of the offered interest rate. This last feature, combined with the price and production update rules, will turn out to be crucial in the dynamics of the model.

The above description defines the basic structure of the Mark I family, but it is of course totally insufficient to code the model, since many additional choices have to be made, leading to several different possible implementations of the model. Here we will use as a baseline model one of the simplest implementation of Mark I, whose description can be found in [34]; the total number of parameters in this version is 10 (but some parameters are actually implicitly fixed from the beginning). We have recoded this basic version and also a slightly different version that we call "Mark I+", which differs on minor details (some that we will specify below) but also on one major aspect: our version of the model *strictly conserves the amount of money in circulation*, i.e. the money in the bank + total firm assets + total households wealth (here savings), in order to avoid – at this stage – any effect due to uncontrolled money creation. A detailed pseudo-code of Mark I+ is provided in Appendix A.

B. State variables

In short (see Appendix A for a complete description), the dynamic evolution of the model is defined by the following state variables. The state of each firm $i = 1 \cdots N_{\rm F}$ is specified by its price $p_i(t)$, the salary it offers $W_i(t)$, its production $Y_i(t)$, its target production $Y_i^T(t)$, its demand $D_i(t)$, its liquidity $\mathcal{L}_i(t)$, its total debt $\mathcal{D}_i^T(t)$. Moreover, each firm is owned by a household and has a list of employees that is dynamically updated. The state of each household $a = 1 \cdots N_{\rm H}$ is specified by its wealth (in the form of savings) $S_a(t)$ and by the firm for which it works (if any).

C. Update rules for prices and production

Among all the micro-rules that any Agent-Based model has to postulate, some seem to be more crucial than others. An important item in Mark I is the behavioural rule for firms adaptation to their economic environment. Instead of the standard, infinite horizon, profit optimizing firm framework (that is both unrealistic and intractable), Mark I postulates a heuristic rule for production $Y_i(t)$ and price $p_i(t)$ update, which reads as follows:

$$\begin{aligned}
&Y_i(t) = D_i(t) \ \&\ p_i(t) > \bar{p}(t) \Rightarrow Y_i^T(t+1) = Y_i(t)[1 + \gamma_y \xi_i(t)] \\
&Y_i(t) = D_i(t) \ \&\ p_i(t) < \bar{p}(t) \Rightarrow p_i(t+1) = p_i(t)[1 + \gamma_p \xi_i(t)] \\
&Y_i(t) > D_i(t) \ \&\ p_i(t) < \bar{p}(t) \Rightarrow Y_i^T(t+1) = Y_i(t)[1 - \gamma_y \xi_i(t)] \\
&Y_i(t) > D_i(t) \ \&\ p_i(t) > \bar{p}(t) \Rightarrow p_i(t+1) = p_i(t)[1 - \gamma_p \xi_i(t)],
\end{aligned} \tag{1}$$

where $D_i(t)$ is the total demand for the goods produced by firm i at time t, and

$$\bar{p}(t) = \frac{\sum_i p_i(t) D_i(t)}{\sum_i D_i(t)} \tag{2}$$

is the average price of sold goods at time t, $\xi_i(t)$ a $U[0,1]$ random variable, independent across firms and across times, and γ_y, γ_p two parameters in $[0,1]$. The quantity $Y_i^T(t)$ is the *target* production at time t, not necessarily the realized one, as described below. These heuristic rules can be interpreted as a plausible *tâtonnement* process of the firms, that attempt to guess their correct production level and price based on the information on the last time step. In spirit, each unit time step might correspond to a quarter, so the order of magnitude of the γ parameters should be a few percent. Note that in the version of Mark I that we consider, wages are fixed to a constant value $W_i(t) \equiv 1$, for all times and all firms.

As we shall see later, the adaptive price/production adjustments described in Eq. (1) leads to two stable attractors (full employment and full unemployment). Which of the two is be reached by the dynamics depends mainly on the level of asymmetry between an upward and downward production adjustments. In Eq. (1), the production adjustment depends on a single parameter γ_y and in this case the system evolves towards a full employment state in the absence of any other constraint. However, as it will become clear in the following section, financial constraints on loans may lead to an effective "weakening" of the upward adjustment, possibly driving the dynamics towards the full unemployment state. As long as such asymmetries between upward and downward production adjustments exist (together with some noise in the price dynamics), the scenario described above and in the following sections is very general and in fact do not depend on details of the update process.

D. Debt and loans

The model further assumes linear productivity, hence the target production corresponds to a target workforce $Y_i^T(t)/\alpha$, where α is a constant coefficient that can always be set to unity (gains in productivity are not considered at this stage). The financial need of the firm is $\max[0, Y_i^T(t) W_i(t) - \mathcal{L}_i(t)]$, where $\mathcal{L}_i(t)$ is the cash available. The total current debt of the firm is $\mathcal{D}_i^T(t)$. The financial fragility of the firm $\ell_i(t)$ is defined in Mark I as the ratio of debt over cash. The offered rate by the banks for the loan is given by:

$$\rho_i(t) = \rho_0 G(\ell_i(t)) \times (1 + \xi_i'(t)), \tag{3}$$

where ρ_0 is the baseline (central bank) interest rate, G is an increasing function (taken to be $G(\ell) = 1 + \tanh(\ell)$ in the reference Mark I and $G(\ell) = 1 + \ln(1+\ell)$ in Mark I+), and ξ' another noise term drawn from a uniform distribution $U[0,1]$. Depending on the rate offered, firms decide to take the full loan they need or only a fraction $F(\rho)$ of it, where $F \leq 1$ is a decreasing function of ρ, called "credit contraction". For example, in the reference Mark I, $F(\rho \leq 5\%) = 1$

and $F(\rho > 5\%) = 0.8$. We have played with the choice of the two functions F, G and the phase transition reported below is in fact robust whenever these functions are reasonable. In Mark I+, we chose a continuous function, such as to avoid built in discontinuities:[4]

$$F(x) = \begin{cases} 1 & \text{if } x < 5\% \\ 1 - \frac{x-5\%}{5\%} & \text{if } 5\% < x < 10\% \\ 0 & \text{if } x > 10\%. \end{cases} \tag{4}$$

The important feature here is that when $F < 1$, the firm does not have enough money to hire the target workforce $Y_i^T(t)$ and is therefore obliged to hire less, or even to start firing in order to match its financial constraints. This financial constraint therefore induces an *asymmetry* in the hiring/firing process: when firms are indebted, hiring will be slowed down by the cost of further loans. As we will see later, this asymmetry is responsible for an abrupt change in the steady state of the economy.

E. Spending budget and bankruptcy

Firms pay salaries to workers and households determine their budget as a fraction c (constant in time and across households) of their total wealth (including the latest salary). Each household then selects M firms at random and sorts them according to their price; it then buys all it can buy from each firm sequentially, from the lower price to the highest price[5]. The budget left-over is added to the savings. Each firm sells a quantity $D_i(t) \leq Y_i(t)$, compute its profits (that includes interests paid on debt), and updates its cash and debt accordingly. Moreover, each firm pays back to the bank a fraction τ of its total debt $\mathcal{D}_i^T(t)$. It also pays dividends to the firm owners if profits are positive. Firms with negative liquidity $\mathcal{L}_i(t) < 0$ go bankrupt. In Mark I+, the cost of the bankruptcy (i.e. $-\mathcal{L}_i(t)$) is spread over healthy firms and on households. Once a firm is bankrupt it is re-initialized in the next time step with the owner's money, to a firm with a price and production equal to their corresponding average values at that moment in time, and zero debt (see Appendix A for more precise statements).

F. Numerical results: a phase transition

When exploring the phase space of Mark I, it soon becomes clear that the baseline interest rate ρ_0 plays a major role. In order not to mix different effects, we remove altogether the noise term ξ in Eq. (3) that affects the actual rate offered to the firms. We find that as long as ρ_0 is smaller than a certain threshold ρ_c, firms are on average below the credit contraction threshold and always manage to have enough loans to pay wages. In this case the economy is stable and after few (~ 100) time steps reaches a stationary state where the unemployment rate is low. If on the other hand the baseline interest rate ρ_0 exceeds a critical value ρ_c, firms cannot afford to take as much loans as they would need to hire (or keep) the desired amount of workers. Surprisingly, this induces a sudden, catastrophic breakdown of the economy. Production collapses to very small values and unemployment sky-rockets. This transition between two states of the economy takes place in both the reference Mark I and in the modified Mark I+; as we shall show in the next section, this transition is actually generic and occurs in simplified models as well. Note in particular that ρ_c is *different* from the value at which $F(x)$ starts decreasing.

The data we show in Fig. 2 corresponds to Mark I+ with parameters $\gamma_p = \gamma_y = 0.1$ and $M = 3$ (see Appendix A for the general parameter setting of the model). While the qualitative behaviour of the model is robust, the details of the transition may change with other parameter settings. For example, smaller values of γ_p, γ_y lead to lower critical thresholds ρ_c (as well as smaller values of M) and to longer equilibration times (T_{eq} scales approximately as $1/\gamma_{y,p}$ for $\rho_0 < \rho_c$). Increasing the size of the economy only affects the magnitude of the fluctuations within one phase leaving the essential features of the transition unchanged. Interestingly, although it is not clear from Fig. 2, the model exhibits oscillatory patterns of the employment rate. The presence of these oscillations can be seen in the frequency domain of the employment rate time series (not shown here), which is essentially characterized by a white noise power spectrum

---

[4] Note however that the arbitrary thresholds (5% and 10%) in Eq. (4) are of little importance and only affect the precise location of the phase transition.

[5] In this sense, the good market is not efficient since the household demand is not necessarily satisfied when M is small. The job market instead, though not efficient due to the presence of unvoluntarely unemployment, is characterized by perfect information since all the workers can contact all the firms until all the open positions are filled.

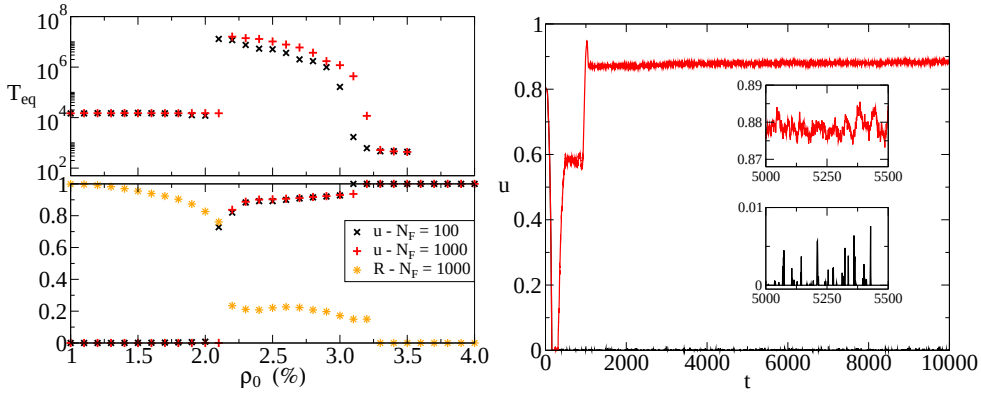

FIG. 2: *Left:* Average unemployment rate u as a function of the interest rate ρ_0 for two system sizes (with $N_{\rm H} = 10N_{\rm F}$ in both cases). The average is over 100 000 time steps discarding the first 50 000 time steps. The phase transition at $\rho_c = 2.1\%$ is of first order with u jumping discontinuously from small values to 1 (intermediate values obtained for $\rho_0 \in [2.1\%, 3.1\%]$ are only due to a much longer equilibration time near the critical point). Note that in the bottom graph we show averages for $T_{eq} = 50\,000$ regardless of ρ_0 while an estimate of the time needed to reach the steady state as a function of ρ_0 is plotted in logarithmic scale in the top graph. In the bottom graph we also show the average value of the asymmetry measure R (see main text) for $N_{\rm F} = 1000$. As one can see R is decreasing with the interest rate up to a point where the asymmetry is strong enough to drive the economy into the high unemployment phase. *Right:* Two trajectories of the unemployment rate with $N_{\rm F} = 1000$ at $\rho_0 = 1.9\%$ and $\rho_0 = 2.5\%$.

with a well defined peak at intermediate frequencies. All these effects will become clearer within the reduced model described in the next section.

Anticipating the results of the next section, we have characterized, as a function of ρ_0, the asymmetry ratio R that characterizes how firms react to the need to hire or to fire. More precisely, we compute the ratio between the *target* number of job creations (resp. destructions) – as a reaction to excess demand (or supply) in the previous time step – to the *realized* number of hires (fires) after financial constraints are met. At each time step we can then extract the average target-to-realized ratios for firms in excess supply and demand respectively. The ratio R (averaged over time) between these two numbers is a proxy of the asymmetry with which firms react to the need to hire or fire. It is clear that increasing ρ_0 hobbles the capacity of firms to hire when they need to, and therefore decreases the ratio R, as seen in Fig. 2, which suggests that the employment collapse is related to this ratio. The results of the next section will fully confirm this scenario.

To sum up, our most salient finding is that Mark I (or Mark I+) has essentially two stationary states, with a first order (discontinuous) transition line separating the two. If the parameters are such that the system lies close to this critical line, then any small modulation of these parameters – such as the noise term that appears in Eq. (3) – will be *amplified* by the proximity of the transition, and lead to interesting boom/bust oscillations, of the kind originally observed in Mark I [34], and perhaps of economic relevance. The question, of course, is how generic this scenario is. We will now show, by studying much simplified versions of Mark I, that this transition is generic, and can indeed be induced by the asymmetry between hiring and firing. We will then progressively enrich our watered-down model (call "Mark 0" below) and see how the qualitative picture that we propose is affected by additional features.

III. HYBRID ABM'S: THE MINIMAL "MARK 0" MODEL

Moving away from the RA framework, Agent-Based Modeling bites the bullet and attempts to represent in details *all* the individual components of the economy (as, for example, in [28, 29]). This might however be counter-productive, at least in the research stage we are still in: keeping too many details is not only computer-time consuming, it may also hobble the understanding of the effects that ABMs attempt to capture. It may well be that some sectors of the economy can be adequately represented in terms of aggregate variables, while discreteness, heterogeneities and interactions are crucial in other sectors. In our attempt to simplify Mark I, we posit that the whole household sector can be represented by aggregate variables: total wealth (again entirely in the form of savings) $S(t)$, total income wage $W(t)$ and total consumption budget $C_B(t)$ (which, as we will emphasize below, is in general larger than the *actual* consumption $C(t)$). We also remove the banking sector and treat the loans in the simplest possible way – see below. While the interest rate is zero in the simplest version, the incentive to hire/fire provided by the interest rate, that was at play in Mark I, will be encoded in a phenomenological way in the update rule for production. The firms,

on the other hand, are kept as individual entities (but the above simplifications will allow us to simulate very large economies, with $N_{\rm F} = 100\,000$ firms or more).

A. Set-up of the model

The minimal version of the Mark 0 model is defined as follows. The salient features are:

- There are $N_{\rm F}$ firms in total and $\mu N_{\rm F}$ households, $\mu > 1$.[6] Each firm $i = 1 \cdots N_{\rm F}$ pays a salary $W_i(t)$ and produces output $Y_i(t)$ by means of a one-to-one technology that uses only labor as an input. Productivity is chosen to be *constant in time* and fixed to $\alpha = 1$. We therefore neglect any productivity shocks in our model; interestingly, crises (when they occur) will be of endogeneous origin. With $\alpha = 1$, $Y_i(t)$ is simply equal to the number of employees of firm i. Hence, the employment rate $\varepsilon(t)$ and unemployment rate $u(t)$ are:

$$\begin{aligned} \varepsilon(t) &= \frac{1}{\mu N_{\rm F}} \sum_i Y_i(t) \ , \\ u(t) &= 1 - \varepsilon(t) \ . \end{aligned} \tag{5}$$

- Households are described by their total accumulated savings $S(t)$ (which at this stage are always non-negative) and by their total wage income $\sum_i W_i(t) Y_i(t)$. At each time step, they set a total consumption budget[7]

$$C_B(t) = c\,[S(t) + \sum_i W_i(t) Y_i(t)] \tag{6}$$

  which is distributed among firms using an intensity of choice model [36]. The demand of goods for firm i is therefore:

$$D_i(t) = \frac{C_B(t)}{p_i(t)} \frac{e^{-\beta p_i(t)/\overline{p}(t)}}{Z(t)} \ , \qquad Z(t) = \sum_i e^{-\beta p_i(t)/\overline{p}(t)} \tag{7}$$

  where β is the price sensitivity parameter determining an exponential dependence of households demand in the price offered by the firm; $\beta = 0$ corresponds to complete price insensitivity and $\beta \to \infty$ means that households select only the firm with the lowest price.[8] The normalization is such that $C_B(t) = \sum_i p_i(t) D_i(t)$, as it should be.

- Firms are described by their price $p_i(t)$, their salary $W_i(t)$, and their production $Y_i(t)$.
  - For simplicity, we fix the salary $W_i(t) \equiv 1$ – an extension that includes wage dynamics is discussed below.
  - For the price, we keep the Mark I price update rule (1), with the average production-weighted price:

$$\overline{p}(t) = \frac{\sum_i p_i(t) Y_i(t)}{\sum_i Y_i(t)} \ . \tag{8}$$

    Note that this price update rule only makes sense if firms anticipate that households are price sensitive, i.e. if $\beta > 0$, which we will assume in the following. Still, the dynamics of the model as defined remains perfectly well-behaved when $\beta = 0$, even if in this limit, the rational behaviour of firms would be to increase their price indefinitely and produce very little.

---

[6] Actually, households are treated as a unique aggregate variable, therefore μ is not a relevant parameter: one can see that its value is irrelevant and one can always set $\mu = 1$ for simplicity. Yet it is useful to think that the aggregate variables represents in an effective way a certain number of individual households, hence we keep the parameter μ explicit in the following.

[7] Of course, one could choose different c's for the fraction of savings and the fraction of wages devoted to spending, or any other non-linear spending schedule.

[8] In this sense, as long as $\beta > 0$ firms compete on prices. An averaged scatter plot of firms profits versus the price offered (not shown here) indeed displays a well-shaped concave profit function.

– For production, we assume that firms are more careful with the way they deal with their workforce than posited in Mark I. Independently of their price level, firms try to adjust their production to the observed demand. When firms want to hire, they open positions on the job market; we assume that the total number of unemployed workers, which is $\mu N_{\rm F} u(t)$, is distributed among firms according to an intensity of choice of model which depends on both the wage offered by the firm[9] and on the same parameter β as it is for Eq. (7); therefore the maximum number of available workers to each firm is:

$$\mu\tilde{u}_i(t) = \frac{e^{\beta W_i(t)/\overline{w}(t)}}{\sum_i e^{\beta W_i(t)/\overline{w}(t)}} \mu N_{\rm F} u(t) \ . \tag{9}$$

where

$$\overline{w}(t) = \frac{\sum_i W_i(t) Y_i(t)}{\sum_i Y_i(t)} \ . \tag{10}$$

In summary, we have

$$\begin{array}{lll} \text{If } Y_i(t) < D_i(t) & \Rightarrow & \left\{ \begin{array}{l} Y_i(t+1) = Y_i(t) + \min\{\eta_+(D_i(t) - Y_i(t)), \mu\tilde{u}_i(t)\} \\ \text{If } p_i(t) < \overline{p}(t) \quad \Rightarrow \quad p_i(t+1) = p_i(t)(1+\gamma_p \xi_i(t)) \\ \text{If } p_i(t) \geq \overline{p}(t) \quad \Rightarrow \quad p_i(t+1) = p_i(t) \end{array} \right. \\ \text{If } Y_i(t) > D_i(t) & \Rightarrow & \left\{ \begin{array}{l} Y_i(t+1) = \max\{Y_i(t) - \eta_-[Y_i(t) - D_i(t)], 0\} \\ \text{If } p_i(t) > \overline{p}(t) \quad \Rightarrow \quad p_i(t+1) = p_i(t)(1-\gamma_p \xi_i(t)) \\ \text{If } p_i(t) \leq \overline{p}(t) \quad \Rightarrow \quad p_i(t+1) = p_i(t) \end{array} \right. \end{array} \tag{11}$$

where $\eta_\pm \in [0,1]$ are what we denote as the hiring/firing propensity of the firms. Note that this rule ensures that there is no overshoot in production; furthermore the max in the second rule is not necessary mathematically when $\eta_- \leq 1$, but we kept it for clarity. Each row of Eq. (11) specifies an adjustment mechanism for output and the individual price. According to this mechanism, there is an increase in output if excess demand is positive, and a decrease in output if excess demand is negative, i.e. if there is excess supply. The propensities to hire/fire $\eta_\pm$ can be seen as the sensitivity of the output change to excess demand/supply. These sensitivities are generally less than unity (i.e. the firm is not adjusting output one to one with excess demand/supply), because there are hiring (firing) costs of different kinds (real costs, time-to-hire, administrative constraints, inertia, etc.). Firing costs generate "labour hoarding", while hiring costs prevents the economy to adapt quickly to excess demand. Hiring and firing costs may not be identical. Therefore the sensitivity to excess demand (or hiring propensity η_+) may be different from the sensitivity to excess supply (or firing propensity η_-). Note that because of the $\min\{D_i(t) - Y_i(t), \mu u(t)\}$ term in Eq. (11), the total production of the model is bounded by $\sum_i Y_i(t) = \mu N_{\rm F} \varepsilon(t) \leq \mu N_{\rm F}$, as it should be because $\varepsilon(t) = 1$ corresponds to full employment and in that case $\sum_i Y_i(t) = \mu N_{\rm F}$. However, in the following we will sometimes (when stated) remove this bound for a better mathematical tractability. This amounts to replace $\min\{D_i(t) - Y_i(t), \mu u(t)\} \to D_i(t) - Y_i(t)$ in Eq. (11) (it corresponds to choosing $\mu = \infty$). Removing the bound corresponds to a situation where labor resources can freely exceed the working population, such that the high employment phase translates into an exponential explosion of the economy output, which is of course unrealistic.

- *Accounting of firms and households.* Each firm $i = 1 \cdots N_{\rm F}$ pays a total wage bill $W_i(t)Y_i(t)$ and has total sales $p_i(t)\min\{Y_i(t), D_i(t)\}$. Moreover, if the profit of the firm $\mathcal{P}_i(t) = p_i(t)\min\{Y_i(t), D_i(t)\} - W_i(t)Y_i(t)$ is positive, the firm pays a dividend $\delta \times \mathcal{P}_i(t)$ to the households.[10]

  Note that if $D_i > Y_i$, the demand for goods of firm i cannot be immediately satisfied, and we assume that in this case households involuntarily save and delay their consumption until the next round (but still using Eq. (6) with the correctly updated savings). The *actual* consumption $C(t)$ (limited by production) is therefore given by:

$$C(t) := \sum_{i=1}^{N_{\rm F}} p_i(t)\min\{Y_i(t), D_i(t)\} \leq C_B(t) = \sum_{i=1}^{N_{\rm F}} p_i(t) D_i(t). \tag{12}$$

---

[9] Since at this stage wages are equal among firms the distribution is uniform. Below we allow firms to update their wage. A higher wage will then translate in the availability of a larger share of unemployed workers in the hiring process.

[10] We have also considered the case where firms distribute a fraction δ of the profits *plus* the reserves. See below.

In summary, the accounting equations for total accumulated savings $S(t)$ and firms' net deposits $\mathcal{E}_i(t)$ – possibly negative – are the following (here $\theta(x \geq 0) = 1$ and $\theta(x < 0) = 0$ is the Heaviside step function):

$$\begin{aligned} &\mathcal{P}_i(t) = p_i(t) \min\{Y_i(t), D_i(t)\} - W_i(t) Y_i(t) \ , \\ &\mathcal{E}_i(t+1) = \mathcal{E}_i(t) + \mathcal{P}_i(t) - \delta \mathcal{P}_i(t) \theta(\mathcal{P}_i(t)) \ , \\ &S(t+1) = S(t) - \sum_i \mathcal{P}_i(t) + \delta \sum_i \mathcal{P}_i(t) \theta(\mathcal{P}_i(t)) \equiv S(t) + \mathcal{I}(t) - C(t) \ , \end{aligned} \tag{13}$$

where $\mathcal{I}(t)$ is the total income of the households (wages plus dividends) and $C(t)$ is the money actually spent by households, which is in general less than their consumption budget $C_B(t)$. This corresponds to *unvoluntarly* households savings whenever production is below demand.

Note that total money $S(t) + \sum_i \mathcal{E}_i(t)$ is clearly conserved since $\Delta S(t) = -\Delta[\sum_i \mathcal{E}_i(t)]$ for cash-flow consistency in a closed economy.

- *Bank accounting.* As mentioned above, we allow the firms' net deposits to become negative, which we interpret as the firm being in need of an immediate extra line of credit. Depending on the financial fragility of the firm (defined below), the bank may or may not agree to restructure the debt and provide this extra credit. If it does, our accounting procedure can be rephrased in the following way. In case of negative net deposits $\mathcal{E}_i(t) < 0$ the bank provides the firm with the extra liquidity needed to pay the wages. The equity of the firm is therefore equal to its deposits when positive, and is close to zero, but still equal to what it needs, when the net deposits is negative. However, when the firm becomes too indebted, the bank will not provide the liquidity needed to pay the wages, leading to negative equity and bankruptcy. From an accounting perspective the matrix balance sheet of our model can be summarized as follows. We assume the bank's equity to be constant in time (let it be 0 for simplicity). The bank holds at the beginning of the simulation a quantity M of currency (issued by the central bank, not modeled here) and therefore has assets equal to M throughout the simulation.

  Households have deposits $S > 0$ while firms have either deposits (if $\mathcal{E}_i > 0$) or liabilities (if $\mathcal{E}_i < 0$). Defining $\mathcal{E}^+ = \sum_i \max(\mathcal{E}_i, 0)$ and $\mathcal{E}^- = -\sum_i \min(\mathcal{E}_i, 0)$, the balance sheet of the banking system is therefore:

$$M + \mathcal{E}^-(t) = S(t) + \mathcal{E}^+(t) \equiv \mathcal{X}(t), \tag{14}$$

  which means that the total amount of deposits $\mathcal{X}$ at any time is equal to initial deposits plus deposits created by the banking system when issuing loans. In this setting reserves M at the bank are kept unchanged when loans are issued (as outlined for example in [35]) and deposits increase (decrease) only when loans are issued (repaid). Correspondingly, the fundamental time-invariant macroeconomic accounting identity

$$S(t) + \mathcal{E}^+(t) - \mathcal{E}^-(t) = S(t) + \sum_i \mathcal{E}_i(t) = M \tag{15}$$

  is obeyed and amounts to our money conserving equation.

- *Financial fragility and bankruptcy resolution.* We measure the indebtement level of a firm through the ratio of (negative) net deposits over payroll (equal, to a good approximation, to total sales):

$$\Phi_i = -\mathcal{E}_i / (W_i Y_i) \tag{16}$$

  which we interpret as a measure of financial fragility. (Implicitly, this assumes that the "real assets" of the firms – not modeled here – are proportional to total payroll). If $\Phi_i(t) < \Theta$, i.e. when the level of debt is not too high compared to the size of the company, the firm is allowed to continue its activity. If on the other hand $\Phi_i(t) \geq \Theta$, the firm defaults.

  When firms exceeds the bankrupcty threshold the default resolution we choose is the following. We first define the set $\mathcal{H}_i$ of financially "healthy" firms that are potential buyers for the defaulted firm i. The condition for this is that $\mathcal{E}_j(t) > \max(-\mathcal{E}_i(t), \Theta Y_j(t) W_j(t))$, meaning that the firm j has a strongly positive net deposits and can take on the debt of i without going under water.

  - With probability $1 - f$, $f \in [0, 1]$ being a new parameter, a firm j is chosen at random in $\mathcal{H}_i$; j transfers to i the needed money to pay the debts, hence $\mathcal{E}_j \to \mathcal{E}_j + \mathcal{E}_i$ (remember that $\mathcal{E}_i$ is negative) and $\mathcal{E}_i \to 0$ after the transaction. Furthermore, we set $p_i = p_j$ and $W_i = W_j$, and the firm i keeps its employees and its current level of production.

– With probability f, or whenever $\mathcal{H}_i = \emptyset$, the firm i is not bailed out, goes bankrupt and its production is set to zero. In this case its debt $\mathcal{E}_i(t) < 0$ is transferred to the households' accumulated savings, $S(t) \to S(t) + \mathcal{E}_i(t)$ in order to keep total money fixed[11].

Hence, when f is large, most of the bankruptcies load weighs on the households, reducing their savings, whereas when f is small, bankruptcies tend to fragilise the firm sector. (The Mark I+ model discussed above corresponds to $f = 1/2$.) It is important to stress that changing the details of the bankruptcy rules while maintaining proper money conservation does not modify the main qualitative message of our paper. The important point here is that the default costs are transferred to households and firms (to ensure money conservation) and have some repercussion on demand (through households accumulated savings) or on firms fragility (through firms net deposits). This can create default avalanches and crises.

- *Firm revival.* A defaulted firm has a finite probability φ per unit time to get revived; when it does so its price is fixed to $p_i(t) = \overline{p}(t)$, its workforce is the available workforce, $Y_i(t) = \mu u(t)$, and its net deposits is the amount needed to pay the wage bill, $\mathcal{E}_i(t) = W_i(t)Y_i(t)$. This liquidity is provided by the households, therefore $S(t) \to S(t) - W_i(t)Y_i(t)$ when the firm is revived, again to ensure total money conservation. Note that during this bankrupt/revival phase, the households' savings $S(t)$ might become negative: if this happens, then we set $S(t) = 0$ and the debt of households is spread over the firms with positive liquidity[12], proportionally to their current value of $\mathcal{E}_i$, again in order to ensure total money conservation and $S(t) \geq 0$.

The above description contains all the details of the definition of the model, however for full clarification a pseudo-code of this minimal Mark 0 model is provided in Appendix B (together with the extensions discussed in Sec. IV). The total number of relevant parameter of Mark 0 is equal to 9: $c, \beta, \gamma_p, \eta_\pm, \delta, \Theta, \varphi, f$ plus the number of firms $N_{\rm F}$. However, most of these parameters end up playing very little role in determining the *qualitative*, long-time aggregate behaviour of the model. Only two quantities play an important role, and turn out to be:

1. the ratio $R = \eta_+/\eta_-$ between η_+ and η_-, which is meant to capture any asymmetry in the hiring/firing process. As noted above (see Fig. 2), a rising interest rate endogeneously leads to such a hiring/firing asymmetry in Mark I and Mark I+. But other sources of asymmetry can also be envisaged: for example, overreaction of the firms to bad news and under-reaction to good news, leading to an over-prudent hiring schedule. Capital inertia can also cause a delay in hiring, whereas firing can be immediate.

2. the default threshold Θ, which controls the ratio between total debt and total circulating currency. In our minimal setting with no banks, it plays the role of a money multiplier. Monetary policy within Mark 0 boils down to setting of the maximum acceptable debt to payroll ratio Θ.

The other parameters change the phase diagram of the model quantitatively but not qualitatively. In order of importance, the most notable ones are f (the redistribution of debt over households or firms upon bankruptcies) and β (the sensitivity to price) – see below.

B. Numerical results & Phase diagram

When running numerical simulations of Mark 0, we find (Fig. 3) that after a transient that can be surprisingly long[13], the unemployment rate settles around a well defined average value, with some fluctuations (except in some cases where endogenous crises appear, see below).

We find the same qualitative phase diagram for all parameters $\beta, \gamma_p, \varphi, \delta, f$. For a given set of parameters, there is a critical value R_c of $R = \eta_+/\eta_-$ separating a full unemployment phase for $R < R_c$ from a phase where some of the labour force is employed for $R > R_c$. Here $R_c \leq 1$ is a value that depends on all other parameters. In Sec. V we

---

[11] One can interpret this by imagining the presence of a bank that collects the deposits of households and lend money to firms, at zero interest rate. If a firm goes bankrupt, the bank looses its loan, which means that its deposits (the households' savings) are reduced.

[12] Again, this can be interpreted by imagining that the firms with positive liquidity deposit their cash in the bank. When the bank needs to provide a loan to a revived firm, or loses money due to a bankrupt, it prefers to take this money from households' deposits, but if these are not available, then it takes the money from firms' deposits.

[13] Think of one time step as a quarter, which seems reasonable for the frequency of price and workforce updates. The equilibration time is then 20 years or so, or even much longer as for the convergence to the 'bad' state in Mark I, see Fig. 2. Albeit studying a very different ABM, similarly long time scales can be observed in the plots shown in [33]. See also the discussion in the conclusion on this point

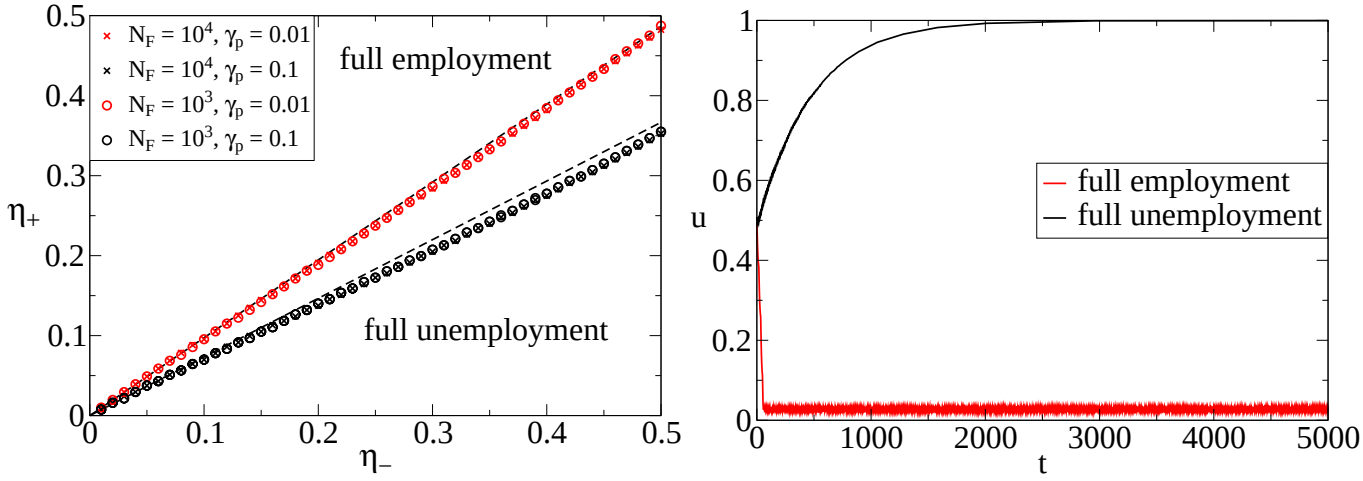

FIG. 3: (*Left*) Phase diagram of the basic Mark 0 model with $\beta = 2$ and $\Theta = \infty$. There are two distinct phases separated by a critical line which depends on the parameters γ_p and β. The dashed lines correspond to the analytical result in Eq. (29) which agrees well with numerical simulations for small values of η_+, η_- where the perturbative method of Section III-C and Appendix C 1 is justified. For simplicity we show here only results with $\beta = 2$ but Eq. (29) is in good agreement with numerical results up to $\beta \sim 4$. (*Right*) Two typical trajectories of $u(t)$ in the two phases ($R = 5/3$ in the full employment line and $R = 3/5$ in the full unemployment line). The other parameters are: $N_{\rm F} = 10\,000$, $\gamma_p = 0.1$, $\Theta = \infty$, $\varphi = 0.1$, $\beta = 2$, $c = 0.5$. $\delta = 0.02$ and $\varphi = 1$.

explain how the phase transition at R_c and other features of the model can be understood by means of approximate analytical calculations.

The transition is particularly abrupt in the limit $\Theta \to \infty$ (no indebtment limit), in which the unemployment rate jumps all the way from 0 to 1 at R_c, see Fig. 3. The figure shows that indeed only the ratio $R = \eta_+/\eta_-$ is relevant, the actual values of $\eta_\pm$ only change the time scale over which the production fluctuates. Moreover, the phase diagram is almost independent of $N_{\rm F}$, which confirms that we are effectively in a limit where the number of firms can be considered to be very large[14].

Interestingly, for finite values of Θ, the phase diagram is more complex and shown in Fig. 4. Besides the full unemployment phase (region 1, which always prevails when $R < R_c$, we find three other different regions for $R > R_c$, that actually survive many extensions of Mark 0 that we have considered (see section IV):

- At very large Θ (region 4, "FE"), the full employment phase persists, although a small value of the unemployment appears in a narrow region around $R \approx R_c$. The width of this region of small unemployment vanishes as Θ increases.
- At very low Θ (region 2, "RU"), one finds persistent "Residual Unemployment" in a large region of $R > R_c$. The unemployment rate decreases continuously with R and Θ but reaches values as large as 0.5 close to $R = R_c$ (see Fig. 4, Bottom Left).
- A very interesting *endogenous crises* phase appears for intermediate values of Θ (region 3, "EC"), where the unemployment rate is most of the time very close to zero, but endogenous crises occur, which manifest themselves as sharp spikes of the unemployment that can reach quite large values. These spikes appear almost periodically, and their frequency and amplitude depend on some of the other parameters of the model, in particular f and β, see Fig. 4, Bottom Right.

The phase diagram in the plane $R - \Theta$ is presented in Fig. 4, together with typical time series of the unemployment rate, for each of the four phases.

We also show in Fig. 5 a trajectory of $u(t)$ in the good phase of the economy (FE, region 4) and zoom in on the small fluctuations of $u(t)$ around its average value. These fluctuations reveal a clear periodic pattern in the low unemployment phase; recall that we had already observed these oscillations within Mark I+. Oscillating patterns

[14] Actually, in the specific case $\Theta = \infty$, even $N_{\rm F} = 1$ provides a similar phase diagram, although the critical R_c slightly depends on $N_{\rm F}$ when this number is small

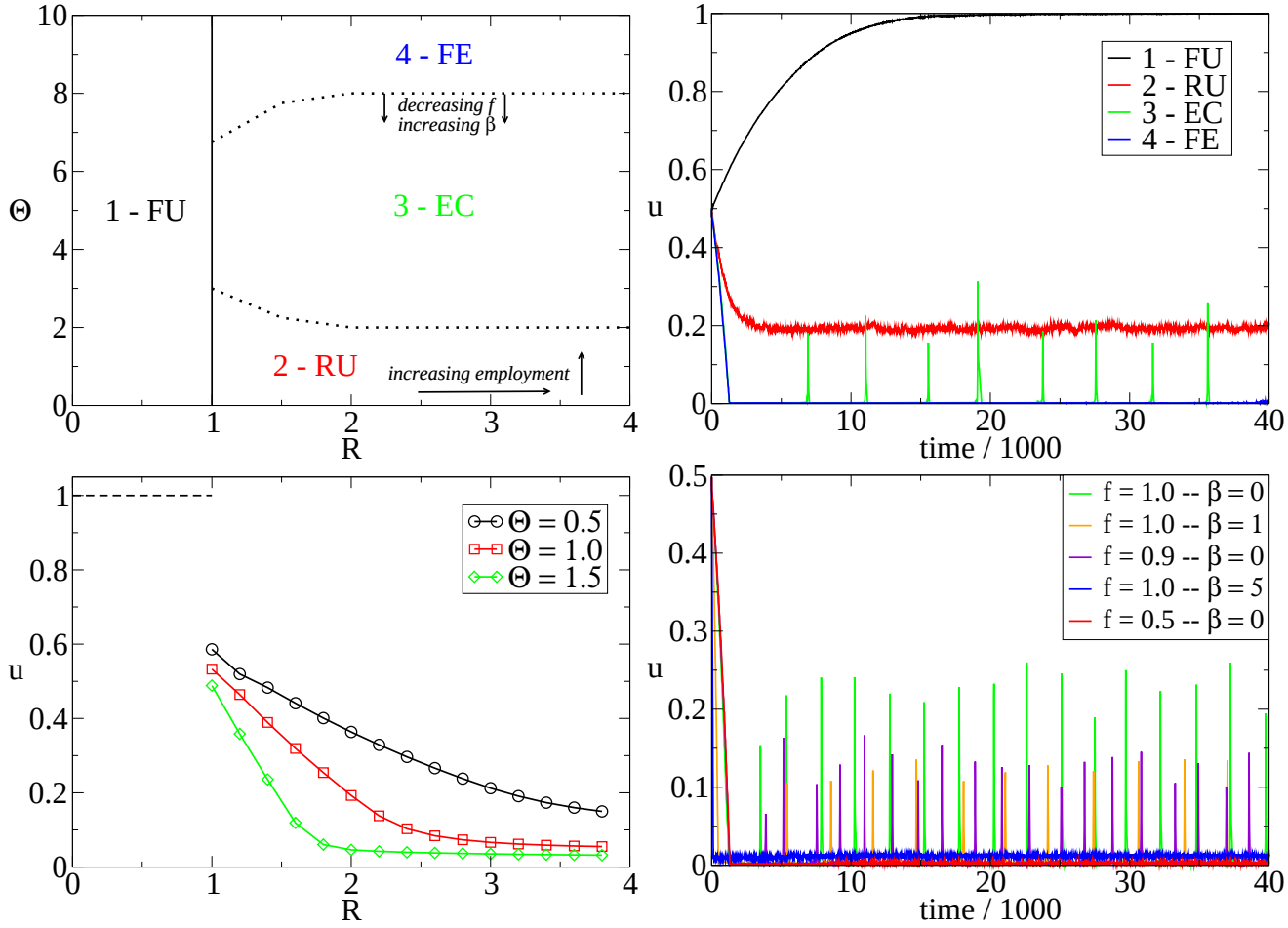

FIG. 4: (*Top Left*) Phase diagram of the basic Mark 0 model in the $R - \Theta$ plane, here with $N_{\rm F} = 5000$, $c = 0.5$, $\gamma_p = 0.05$, $\delta = 0.02$, $\varphi = 0.1$. Here we keep the 'extreme' cases $\beta = 0$ and $f = 1$ since the dependence of phase boundaries on β and on f (here reported schematically) is shown in detail in Fig. III C 3. The other parameters are irrelevant. There are four distinct phases separated by critical lines. (*Top Right*) Typical time series of $u(t)$ for each of the phases. (*Bottom Left*) Stationary value of the unemployment rate as a function of R for different values of Θ in phase 2. (*Bottom Right*) Typical trajectories of $u(t)$ in the Endogenous Crisis phase (region 3) for different values of f and β, the other parameters are kept fixed. Note that increasing β or decreasing f lead to small crisis amplitudes.

(perhaps related to the so-called business cycle) often appear in simplified first order differential models of the macroeconomy; one of the best known examples is provided by the Goodwin model [37, 38], which is akin to a predator-prey model where these oscillations are well known. But these oscillations do also show up in other ABMs, see [31, 33]. Note that these fluctuations/oscillations around equilibrium *do not regress* when the number of firms get larger. We have simulated the model with $N_{\rm F} = 10\,000$ firms and $N_{\rm F} = 1\,000\,000$ firms with nearly identical amplitude and frequencies for these fluctuations. We will offer some insight on the origin of these oscillations in Sec. V. Note that we do not expect such oscillations to remain so regular in the real economy, in particular because exogeneous shocks are absent from our model and because we assume that all firms are characterised by the very same adaptation time scale γ_p.

C. Qualitative interpretation. Position of the phase boundaries

An important quantity that characterizes the behavior of the model in the "good" phase of the economy (i.e. for $R > R_c$) is the global leverage (debt-to-equity) ratio k:

$$k(t) = \frac{\mathcal{E}^-(t)}{S(t) + \mathcal{E}^+(t)} = 1 - \frac{M}{S(t) + \mathcal{E}^+(t)} \tag{17}$$

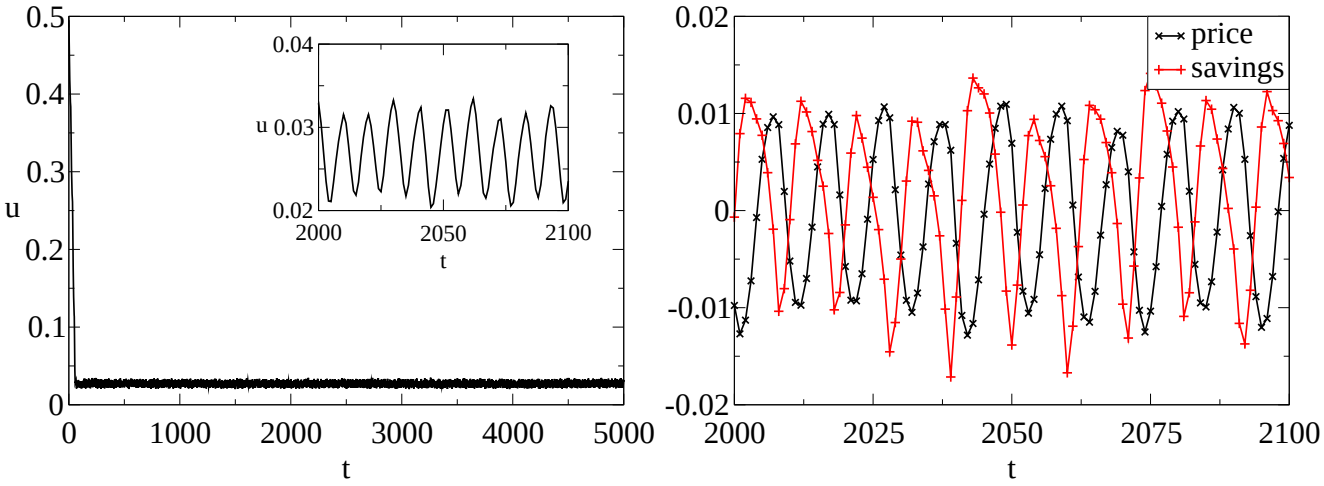

FIG. 5: *Left*: Typical time evolution of the unemployment u, starting from an initial condition $u = 0.5$, for the basic Mark 0 model in the Full Employment phase (region 4). The trajectory leading to a "good" state of the economy is obtained for $\eta_+ = 0.5$ and $\eta_- = 0.3$. Note the clear endogenous "business cycles" that appear in that case. These runs are performed with $N_F = 10\,000$ firms, $c = 0.5$, $\beta = 2$, $\gamma_p = 0.1$, $\varphi = 0.1$, $\Theta = 5$, $\delta = 0.02$.
Right: Oscillations for the average price and the average accumulated savings per household (each shifted by their average values for clarity) for the same run as in the *Left* figure. When prices are low, savings increase, while when prices are high, savings decrease. See Sec. V for a more detailed description.

where, due to money conservation, $k \leq 1$. The good state of the economy is characterized by a large average value of k reflecting the natural tendency of the economy towards indebtment, the level of which being controlled in Mark 0 by the parameter Θ (the average value of k increases with Θ).[15] Interestingly, in regions 2-RU and 4-FE $k(t)$ reaches a stationary state, whereas in region 3-EC its dynamics is characterized by an intermittent behavior corresponding to the appearance of endogenous crises during which indebtment is released through bankruptcies.

1. The EC phase

This phenomenology can be qualitatively explained by the dynamics of the distribution of firms fragilities Φ_i. For $R > R_c$, firms overemploy and make on average negative profits, which means that the Φ_i's are on average drifting towards the bankruptcy threshold Θ.

- When Θ is small enough (i.e. in region 2-RU) the drift is continuously compensated by the reinitialization of bankrupted firms and the fragility distribution reaches a stationary state.
- For intermediate values of Θ, however, (i.e. in region 3-EC) the number of bankruptcies per unit of time becomes intermittent. Firms fragilities now collectively drift towards the bankruptcy threshold; as soon as firms with higher fragilities reach Θ, bankruptcies start to occur. Since for f large enough, bankruptcies are mostly financed by households, demand starts falling which has the effect of increasing further the negative drift. This feedback mechanism gives rise to an avalanche of bankruptcies after which most of the firms are reinitialized with positive liquidities. This mechanism has the effect of *synchronizing* the fragilities of the firms, therefore leading to cyclical waves of bankruptcies, corresponding to the unemployment spikes showed in Fig. 4. The distribution of firms fragilities does not reach a stationary state in this case.
- When $\Theta \gg 1$ (i.e. region 4-FE), households are wealthy enough to absorb the bankruptcy cost without pushing the demand of goods below the maximum level of production reached by the economy.[16] Hence, the economy settles down to a full employment phase with a constant (small) rate of bankruptcies.

---

[15] In a further version of the model, a central bank will be in charge of controlling k through a proper monetary policy.
[16] The amount of money circulating in the economy increases with Θ and for $R > R_c$ it is largely channelled to households savings since firms have on average negative liquidities.

The above interpretation is supported by a simple one-dimensional random walk model for the firms assets, with a drift that slef-consistently depends on the number of firms that fail. This highly simplified model accurately reproduces the above phenomenology, and is amenable to a full analytical solution, which will be published separately.

The existence of the EC phase is a genuinely surprising outcome of the model, which was not put by hand from the outset (see our discussion in the first lines of the Introduction section above). Crises occur there not as a result of sweeping parameters through a phase transition (as is the case, we argued, of Mark I with a time dependent interest rate) but purely as a result of the own dynamics of the system.

2. The role of β

The dependence of the phase boundaries on the different parameters is in general quite intuitive. The dependence of aggregate variables on β, for example, is interesting: everything else being kept equal, we find that increasing β (i.e. increasing the price selectivity of buyers), increases the level of unemployment (see inset of Fig. 6) and the amplitude of the fluctuations around the average value (a similar effect was noted in [29]). Increasing γ_p increases the dispersion of prices around the average value and is thus similar to increasing β.

Increasing β has also, within the present setting of Mark 0, some counter-intuitive effects: it increases both the average price compared to wages and the profits of firms, hence stabilizing the FE phase and shifting its boundary with EC to lower values of Θ and R_c (see Fig. 6 for the amplitude of region 3-EC as a function of β. This effect can be understood by considering the demand-production gap $G_i = D_i - Y_i$ as a function of the price difference $\delta p_i = p_i - \bar{p}$. For a fixed value of G_i the rules for price and production updates are independent of β; however, the response of the demand to a price change is stronger for higher values of β: for small values of δp_i one finds approximately $G_i \approx -C(\beta)\delta p_i$, where $C(\beta) > 0$ is increasing with β. As a consequence, the absolute value of the gap, $|G_i|$ for a given value of $|\delta p_i|$ is on average increasing with β. Therefore, the total amount of unsold goods $\sum_i \max[Y_i - D_i, 0]$ and households accumulated savings S also increase with β: households involuntarily save more when they are more selective on prices. An increased amount of savings is in turn responsible for the average price increase while the households' higher wealth expands the FE region by shifting its boundary to lower values of Θ. Note, however, the effect of β on the average price is numerically very small and depends sensitively on the precise consumption rule. For example, if we insist that households fully spend their consumption budget $C_B(t)$ by looking for available products at a higher price (as in Mark I), then the above effect disappears (see Section V).

3. The role of f

A potentially more relevant discussion concerns the effect of the bankruptcies on the financial health of the firm sector. Decreasing f (i.e. the financial load taken up by households when bankruptcies occur) also stabilize, as expected, the full employment phase. Such a stabilization can also be achieved by distributing a fraction δ^+ of the profit plus the total positive liquidity of the firms (instead of a fraction δ of the profits only), which has the obvious effect of supporting the demand. In fact, we find that as soon as $f \leq 0.81$ or $\delta^+ \geq 0.25$, the Endogenous Crisis phase disappears and is replaced by a continuous cross-over between the Residual Unemployment phase and the Full Employment phase (see Fig. 6).

D. Intermediate conclusion

The main message of the present section is that in spite of many simplifications, and across a broad range of parameters, the phase transition observed in Mark I as a function of the baseline interest rate is present in Mark 0 as well. We find that these macroeconomic ABMs generically display two very different phases – high demand/low unemployment vs. low demand/high unemployment, with a boundary between the two that is essentially controlled by the asymmetry between the hiring and firing propensity of the firms (compare Figs. 2 & 3).

Moreover, in the Mark 0 model there is an additional splitting of the low unemployment phase in several regions characterized by a different dynamical behaviour of the unemployment rate, depending on the level of debt firms can accumulate before being forced into bankruptcy. We find in particular an intermediate debt region where endogenous crises appear, characterized by acute unemployment spikes. This Endogenous Crisis phase disappears when households are spared from the financial load of bankruptcies and/or when capital does not accumulate within firms, but is transferred to households through dividends. Clearly, these different phases will coexist if the model parameters themselves evolve with time, which one should expect in a more realistic version of the model.

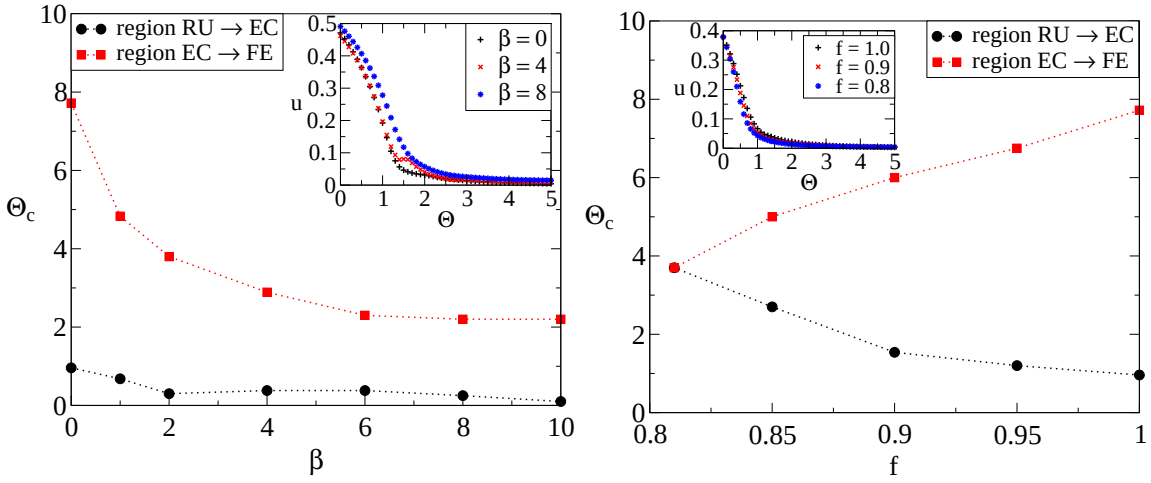

FIG. 6: Phase boundary location Θ_c as a function of the price sensitivity β (left) and of the debt share parameter f (right). Circles correspond to the location of the boundary between regions 2-RU and 3-EC while squares correspond to the boundary between regions 3-EC and 4-FE. In the inset of both figures we plot the average unemployment rate as a function of Θ for different values of β and f. For $f \leq 0.81$ region 4-EC disappears and the average value of the unemployment continuously goes to 0 without the appearance of endogenous crises. Increasing β shrinks the amplitude of region 3-EC which however remains finite for $\beta \gg 1$. As a criterion for being in region 3-EC we require that the amplitude of the crises $\max(u) - \min(u)$ stays above 5%. Parameters are: $N = 5000$, $R = 3$ ($\eta_m = 0.1$), $\delta = 0.02$, $\varphi = 0.1$, $\gamma_p = 0.05$.

We now turn to the first extension of Mark 0 allowing for a wage dynamics that lead to long term inflation or deflation, absent in the above version of the model. Finally, several other extensions of the model will be briefly discussed and presented in a separate publication.

IV. EXTENSION OF MARK 0: WAGE UPDATE

As emphasized above, the Mark I+ and Mark 0 models investigated up to now both reveal a generic phase transition between a "good" and a "bad" state of the economy. However, many features are clearly missing to make these models convincing – setting up a full-blown, realistic macroeconomic Agent-Based Model is of course a long and thorny endeavour which is precisely what we want to avoid at this stage, focusing instead on simple mechanisms. Still, it is interesting to progressively enrich these simplified models not only to test for robustness of our phase diagram but also to investigate new effects that are of economic significance. We consider here wage dynamics, which is obviously an important ingredient in reality. Wage dynamics leads to some relevant effects, in particular the appearance of inflation. Wage dynamics is indeed an item missing from both the basic Mark 0 and Mark I models considered above, which assume fixed wages across time and across firms. Clearly, the ability to modulate the wages is complementary to deciding whether to hire or to fire, and should play a central role in the trajectory of the economy as well as in determining inflation rates.

Introducing wages in Mark 0 again involves a number of arbitrary assumptions and choices. Here, we follow (in spirit) the choices made in Mark I for price and production update, and propose that at each time step firm i updates its wage as:

$$\begin{aligned} W_i^T(t+1) &= W_i(t)[1+\gamma_w \varepsilon \xi_i'(t)] \quad \text{if} \quad \begin{cases} Y_i(t) & < D_i(t) \\ \mathcal{P}_i(t) & > 0 \end{cases} \\ W_i(t+1) &= W_i(t)[1-\gamma_w u \xi_i'(t)] \quad \text{if} \quad \begin{cases} Y_i(t) & > D_i(t) \\ \mathcal{P}_i(t) & < 0 \end{cases} \end{aligned} \tag{18}$$

where $u = 1 - \varepsilon$ is the unemployment rate and γ_w a certain parameter; $\mathcal{P}_i(t) = \min(D_i(t), Y_i(t))p_i(t) - W_i(t)Y_i(t)$ is the profit of the firm at time t and $\xi_i'(t)$ an independent $U[0,1]$ random variable. If $W_i^T(t+1)$ is such that the profit of firm i at time t with this amount of wages would have been negative, $W_i(t+1)$ is chosen to be exactly at the equilibrium point where $\mathcal{P}_i = 0$, hence $W_i(t+1) = \min(D_i(t), Y_i(t))p_i(t)/Y_i(t)$; otherwise $W_i(t+1) = W_i^T(t+1)$.

The above rules are intuitive: if a firm makes a profit and it has a large demand for its good, it will increase the pay of its workers. The pay rise is expected to be larger if unemployment is low (i.e. if ε is large) because pressure

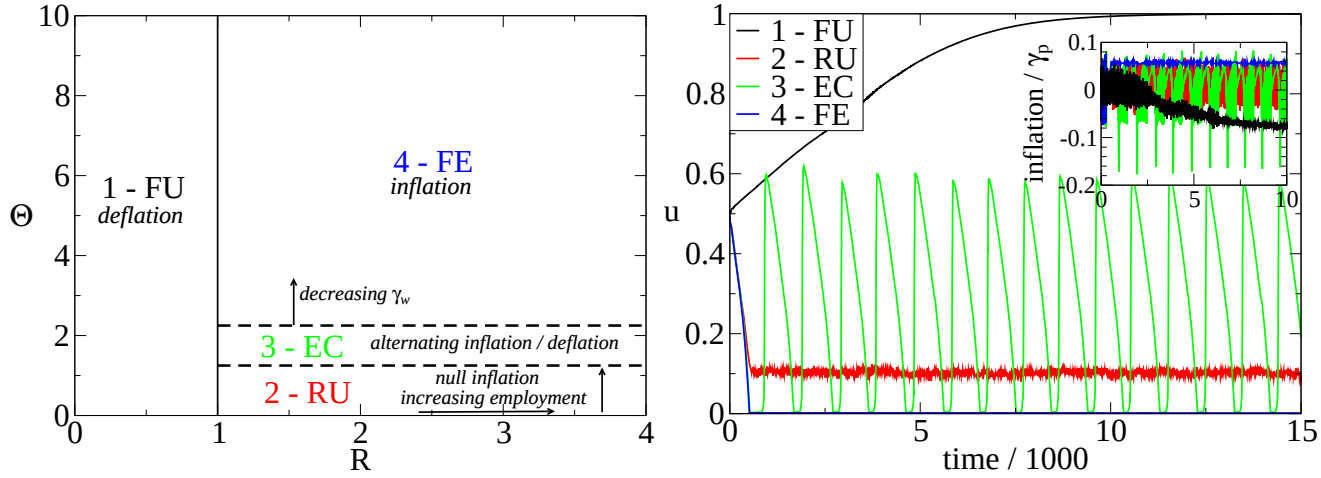

FIG. 7: (*Left*) Phase diagram in the $R-\Theta$ plane of the extended Mark 0 model with wage update with $\gamma_w = \gamma_p$. The parameters of the basic model are the same as in Fig. 4: $N_F = 5000$, $c = 0.5$, $\gamma_p = 0.05$, $\delta = 0.02$, $\varphi = 0.1$, $f = 1$, $\beta = 0$. As for the basic model, there are four distinct phases separated by critical lines. Wage update brings in inflation in the Full Employment phase, and deflation in the full unemployment phase. Endogenous crises are characterized by alternating cycles of inflation and deflation. Interestingly, we find that the location of phase boundaries is in this case almost untouched by changes in both β and f; the other parameters are irrelevant. Note that the wage update strongly stabilizes the full employment phase. (*Right*) A typical trajectory of $u(t)$ for each of the phases. In the inset, the price dynamics is shown, displaying inflation and deflation.

on salaries is high. Conversely, if the firm makes a loss and has a low demand for its good, it will reduce the wages. This reduction is larger when unemployment is high because pressure on salaries is low. In all other cases, wages are not updated.

When a firm is revived from bankruptcy (with probability φ per unit time), its wage level is set to the production weighted average wage of all firms in activity.

The parameters $\gamma_{p,w}$ allow us to simulate different price/wage update timescales. In the following we set $\gamma_p = 0.05$ and $\gamma_w = z\gamma_p$ with $z \in [0, 1]$. The case $z = 0$ clearly corresponds to removing completely the wage update rule, such that the basic version of Mark 0 is recovered. The extended version of Mark 0 that we consider below is therefore characterized by *single additional parameter* $\gamma_w = z\gamma_p$, describing the frequency of wage updates.

A. Results: variable wages and the appearance of inflation

In our *money conserving* toy economies[17], a stationary inflation rate different from zero is possible as long as the ratio $\overline{p}(t)/\overline{W}(t)$ fluctuates around a steady value. In absence of wage update, we have a fixed $W \equiv 1$ and inflation is therefore impossible. The main effect induced by wage dynamics is therefore the possibility of inflation.

Using the wage update rules defined below, we found that the average inflation rate depends on parameters such as the households propensity to consume c and the price/wage adjustment parameters $\gamma_{p,w}$. Most interestingly, we observe a strong dependence of the inflation rate upon the bankruptcy threshold Θ, with large Θ's triggering high inflation and low Θ's corresponding to zero inflation. For intermediate Θ's, periods of inflation and deflation may alternate and the model displays interesting instabilities.

We now analyze the influence of wage adjustments on the phase transitions discussed in the previous sections. The phase diagram for $z = 1$ ($\gamma_w = \gamma_p$) is reported in Fig. 7. The phenomenology that we find is again very similar to the simple Mark 0 without wage update, except for inflation. The most interesting effect is the appearance of inflation in the "good" phases of the economy, and deflation in the "bad" phases, as shown in Fig. 7. When $\Theta \gg 1$, we find again a first order critical boundary at $R = R_c$ that separates a high unemployment phase (with deflation) from a low unemployment phase (with inflation). For $R > R_c$ we see again two additional phases: "EC", with endogenous crises and, correspondingly, alternating periods of inflation and deflation but stable prices on the long run, and "RU", for small Θ, where there is no inflation but a substantial residual unemployment rate (see Fig. 7).

[17] Recall that the physical money is conserved in the model, but virtual money creation is still possible through firms' indebtement.

The appearance of endogenous crises is consistent with what discussed in section III B and is related to situations in which the debt-to-savings ratio $k(t)$ grows faster than prices. From this point of view, increasing γ_w allows firms to better adapt wages (and thus prices) and to absorb the indebtment through inflation; region EC indeed shrinks when increasing γ_w. Note that relating the parameters $\gamma_{p,w}$ to the flexibility of the labor and goods markets is not straightforward. In this sense, it would be instead useful to study the effects of *asymmetric* upward/downward wage and price flexibilities (for example by defining different $\gamma^{\pm}_{p,w}$) in order to understand whether improving the flexibility of the labor and goods markets destabilizes the economy [40–42]. If, however, one considers the ratio $z = \gamma_w/\gamma_p$ as an indicator of wages flexibility (relative to prices flexibility), our results suggests that higher labor flexibility has a stabilizing effect.

We also observe that the oscillatory pattern found for the basic model persists as long as $\gamma_w \ll \gamma_p$, i.e. when wage updates are much less frequent that price updates. The power spectrum of the model for $R > R_c$ and $\Theta \gg 1$ is still characterized by the appearance of a peak (roughly corresponding to a period of 7 time steps). Interestingly, we find that upon increasing the ratio z, the peak in the frequency spectrum disappears but not in a monotonous fashion; intermediate values of z give rise to even more pronounced oscillations than for $z = 0$, before these oscillations disappear for $z > z_c \approx 0.25$.

In conclusion, the comparison between Figs. 7 and 4 demonstrates the robustness of our phase diagram against changes; introducing wages is a rather drastic modification since it allows inflation to set in, but still does not affect the phase transition at R_c, nor the overall topology of the phase diagram, which confirms its relevance. Interestingly, inflation is present in the good phases of the economy and deflation in the bad phases. Our analytical understanding of these effects is however still poor; we feel it would be important to bolster the above numerical results by solving simpler "toy models" as we do for the basic Mark 0 (see section V and Appendix C 1).

B. Other extensions and policy experiments

One of the final goal of our study is to have a prototype framework where one is able to run meaningful policy experiments. In order to do that the model described so far is lacking a number of important ingredients, namely a central bank exogenously setting interest rates level and the amount of money in circulation. This is the project on which we are currently pursuing, with encouraging results [43]. Still, even within the simplistic framework of Mark 0, one can envisage an interesting prototype policy experiment which consists in allowing the "central bank" to temporarily increase the bankruptcy threshold Θ in times of high unemployment. Fig. 8 shows an example of this: the economy is, in its normal functioning mode, in the EC (region 3) of the phase diagram, with $R = 2$ and (say) $\Theta = 2$. This leads in general to a rather low unemployment rate u but, as repeatedly emphasized above, this is interrupted by acute endogenous crises. The central bank then decides that whenever u exceeds 10%, its monetary policy becomes accommodating, and amounts to raising Θ from its normal value 2 to – say – $\Theta = 10$. As shown in Fig. 8, this allows the bank to partially contain the unemployment bursts. However, quite interestingly, it also increases the crisis frequency, as if it did not allow the economy to fully release the accumulated stress. We expect that this phenomenology will survive in a more realistic framework.

We have also explored other potentially interesting extensions of Mark 0, for example adding trust or confidence, that may appear and disappear on time scales much shorter than the evolution time scale of any "true" economic factor, and can lead to market instabilities and crises (see e.g. [5, 39, 44]). There are again many ways to model the potentially destabilizing feedback of confidence. One of the most important channel is the loss of confidence induced by raising unemployment, that increases the saving propensity of households and reduces the demand. The simplest way to encode this in Mark 0 is to let the "c" parameter, that determines the fraction of wages and accumulated savings that is devoted to consumption, be an increasing function of the employment rate $\varepsilon = 1 - u$. We indeed find that the confidence feedback loop can again induce purely endogenous swings of economic activity. Similarly, a strong dependence of c on the recent inflation can induce instabilities [43].

V. ANALYTICAL DESCRIPTION

We attempt here to describe analytically some aspects of the dynamics of Mark 0 in its simplest version, namely without bankruptcies ($\Theta = \infty$ for which φ and f become irrelevant), with $\beta \geq 0$, fixed wages $W = 1$, and no dividends ($\delta = 0$). For simplicity, we also fix $\mu = 1$ and $c = 1/2$. The only relevant parameters are therefore β, γ_p and $\eta_{\pm}$. The

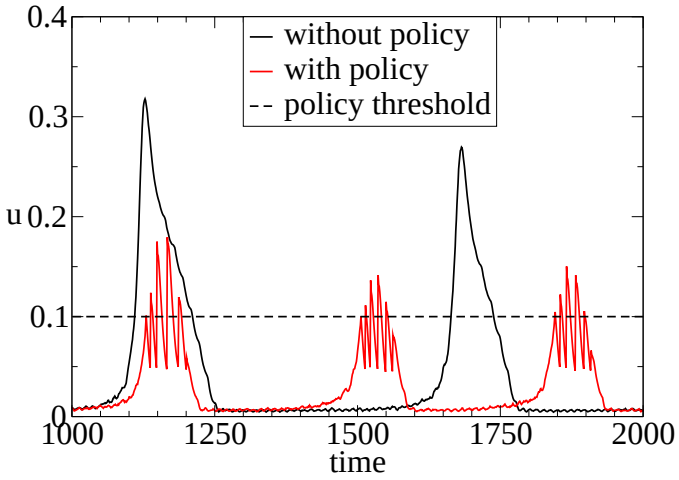

FIG. 8: Here we show an example of a "toy" policy experiment in the Mark 0 model without wage update (i.e $\gamma_w = 0$). We first run a simulation (the "without policy" line) with a constant value $\Theta = 2$ lying in the 3-EC region of Fig. 3. We then run the same simulation with a prototype central bank which increases Θ to 10 as long as u exceeds the threshold of 10%. Note that unemployment is partially contained, but the crisis frequency concomitantly increases. The other parameters values are: $N = 10\,000$, $R = 2$, $\gamma_p = 0.05$, $\varphi = 0.1$, $\delta = 0.02$, $f = 1$.

equations of motion of this very minimal model are:

$$
\begin{aligned}
&\text{If } Y_i(t) < D_i(t) \quad \Rightarrow \quad \left\{ \begin{array}{l} Y_i(t+1) = Y_i(t) + \min\{\eta_+[D_i(t) - Y_i(t)], 1 - \overline{Y}(t)\} \\ \text{If } p_i(t) < \overline{p}(t) \quad \Rightarrow \quad p_i(t+1) = p_i(t)(1 + \gamma_p \xi_i(t)) \\ \text{If } p_i(t) \geq \overline{p}(t) \quad \Rightarrow \quad p_i(t+1) = p_i(t) \end{array} \right. \\
&\text{If } Y_i(t) > D_i(t) \quad \Rightarrow \quad \left\{ \begin{array}{l} Y_i(t+1) = \max\{Y_i(t) - \eta_-[Y_i(t) - D_i(t)], 0\} \\ \text{If } p_i(t) > \overline{p}(t) \quad \Rightarrow \quad p_i(t+1) = p_i(t)(1 - \gamma_p \xi_i(t)) \\ \text{If } p_i(t) \leq \overline{p}(t) \quad \Rightarrow \quad p_i(t+1) = p_i(t) \end{array} \right. \\
&D_i(t) = \frac{c}{p_i(t)} \frac{e^{-\beta(p_i(t) - \overline{p}(t))}}{Z(t)} [\max\{s(t), 0\} + \overline{Y}(t)], \qquad Z(t) := \sum_i e^{-\beta(p_i(t) - \overline{p}(t))} \\
&\mathcal{E}_i(t+1) = \mathcal{E}_i(t) - Y_i(t) + p_i(t) \min\{Y_i(t), D_i(t)\} \\
&s(t) = M_0 - \overline{\mathcal{E}}(t) \ .
\end{aligned}
\tag{19}
$$

Here M_0 is the total money in circulation, whose precise value is irrelevant for this discussion, and $s = S/N_{\mathrm{F}}$ are the savings per agent. Overlines denote an average over firms, which is flat for $\overline{Y} = N_{\mathrm{F}}^{-1} \sum_i Y_i(t)$ and $\overline{\mathcal{E}} = N_{\mathrm{F}}^{-1} \sum_i \mathcal{E}_i(t)$ while it is weighted by production for $\overline{p}$, see Eq. (8). Note that the basic variables here are $\{p_i, Y_i, \mathcal{E}_i\}$, all the other quantities are deduced from these ones.

In the high employment phase, the model admits a stationary state with $\overline{Y}_{st} \sim 1$. To show this, let us focus on the case where γ_p is very small, but not exactly zero, otherwise of course the price dynamics is frozen. The stationary state is attained after a transient of duration $\sim 1/\gamma_p$, and in the stationary state fluctuations between firms are very small in such a way that $p_i \sim \overline{p}$ and $Y_i \sim \overline{Y}$. A stationary state of Eq. (19) has $Y_i(t+1) = Y_i(t)$ and therefore $Y_i = D_i$, hence $Y_i = D_i = \overline{Y}$. Furthermore, from $\mathcal{E}_i(t+1) = \mathcal{E}_i(t)$ we deduce that $p_i = \overline{p} = 1$. Finally, we have $D_i = \overline{Y} = (s + \overline{Y})/2$, which gives $s = \overline{Y}$ and $\mathcal{E} = M_0 - \overline{Y}$. One obtains therefore a continuum of stationary states, because the production $\overline{Y}$ or equivalently the employment are not determined by requiring stationarity. However, we will show in the following that these equilibria become unstable as soon as fluctuations are taken into account ($\gamma_p > 0$). We will see that fluctuations can induce either an exponentially fast decrease of $\overline{Y}$ towards zero (corresponding to the full unemployment phase), or an exponentially fast grow of $\overline{Y}$, which is therefore only cutoff by the requirement $\overline{Y} \leq \mu = 1$, corresponding to full employment. It is therefore natural to choose the stationary state with $\overline{Y} = 1$ as a reference, and study the effect of fluctuations around this state. We therefore consider the stationary state with $Y_i = D_i = p_i = 1$, and $\mathcal{E}_i = M_0 - 1$ in such a way that $s = \overline{Y} = 1$.

For analytical purposes, we will focus below on the limit in which $\eta_\pm, \gamma_p \to 0$, in such a way that we can expand around the high employment stationary state and obtain results for the phase transition point. Numerically, the fact that γ_p and/or $\eta_\pm$ are small does not change the qualitative behavior of the model.

In order to consider small fluctuations around the high employment stationary state, we define the following variables:

$$\begin{aligned} Y_i(t) &= 1 - \gamma_p \zeta_i(t) \ , \\ p_i(t) &= e^{\gamma_p \lambda_i(t)} \ , \\ \mathcal{E}_i(t) &= M_0 - 1 - \gamma_p \alpha_i(t) \ . \end{aligned} \tag{20}$$

Note that in the following, overlines over ζ, α, λ always denote flat averages over firms. Using $p_i(t) = 1 + \gamma_p \lambda_i(t) + (\gamma_p \lambda_i(t))^2/2$ for $\gamma_p \to 0$, one finds that, to order γ_p^2:

$$\frac{e^{-\beta p_i(t)}}{Z(t)} \approx 1 - \beta\gamma_p(\lambda_i(t) - \overline{\lambda}(t)) - \frac{1}{2}\beta\gamma_p^2\Big[(1-\beta)(\lambda_i^2(t) - \overline{\lambda^2}(t)) + 2\beta\overline{\lambda}(t)(\lambda_i(t) - \overline{\lambda}(t))\Big]. \tag{21}$$

Since we will find later that $\overline{\lambda}(t)$ is itself of order γ_p, we will drop the part of the last term in the above expression, which is $O(\gamma_p^3)$, i.e.

$$\frac{e^{-\beta p_i(t)}}{Z(t)} \approx 1 - \beta\gamma_p(\lambda_i(t) - \overline{\lambda}(t)) - \frac{1}{2}\beta(1-\beta)\gamma_p^2\lambda_i^2(t). \tag{22}$$

A. Stability of the high employment phase

To study the stability of the high employment phase, we make two further simplifications.

1. We neglect the fluctuations of the savings and fix $\alpha_i \equiv 0$. This is justified if the employment rate varies slowly over the time scale $\tau_c = -1/\ln(1-c)$ that characterizes the dynamics of the savings. But, as we shall show below, the dynamics of employment becomes very slow in the vicinity of the phase transition, hence this assumption seems justified.

2. In order to proceed analytically, we also consider the limit $\eta_+, \eta_- \to 0$ with $\eta_+ = R\eta_-$ and R of order one. In fact, we will even assume below that $\eta_\pm \ll \gamma_p$. Then, it is not difficult to see that ζ_i is a random variable of order η_+ and is therefore also very small. We define $\zeta_i = \eta_+ z_i$. We believe that these approximations are actually quite accurate, as is confirmed by the comparison of the theoretical transition line with numerical data (see Fig. 3).

With these further simplifications (i.e. setting $\alpha_i = 0$ and assuming that γ_p, $\eta_\pm$ are small and $z_i = O(1)$), we have, together with Eq. (20):

$$\begin{aligned} D_i(t) &= 1 - \gamma_p(\lambda_i(t) + \beta(\lambda_i(t) - \overline{\lambda}(t))) + \frac{1}{2}\gamma_p^2\lambda_i^2(1 + \beta + \beta^2) + O(\gamma_p\eta_+) + O(\gamma_p^3) \ , \\ \overline{p}(t) &= 1 + \gamma_p\overline{\lambda}(t) + \frac{1}{2}\gamma_p^2\overline{\lambda^2}(t) + O(\gamma_p\eta_+) + O(\gamma_p^3) \ , \end{aligned} \tag{23}$$

If, as announced above we further assume that $\eta_+ \ll \gamma_p$, it is justified to keep the terms of order γ_p^2 while neglecting the terms of order $\gamma_p\eta_+$ and γ_p^3.

We also note that if $|\lambda_i|$ is bounded by a quantity of order 1 (which we will find below), then the condition $\lambda_i(t) + \beta(\lambda_i(t) - \overline{\lambda}(t))) - \frac{1}{2}\gamma_p\lambda_i(t)^2 > 0$ is equivalent to $\lambda_i(t) + \beta(\lambda_i(t) - \overline{\lambda}(t))) > 0$. Then the first two lines of Eq. (19) become

$$\begin{aligned} &\text{If } \lambda_i(t) < \frac{\beta}{1+\beta}\overline{\lambda}(t) \quad \Rightarrow \quad \begin{cases} z_i(t+1) = z_i(t) - \min\{-\lambda_i(t) - \beta(\lambda_i(t) - \overline{\lambda}(t)) + \frac{1}{2}\gamma_p\hat{\beta}\lambda_i^2, \overline{z}(t)\} \\ \lambda_i(t+1) = \lambda_i(t) + \xi_i(t) - \frac{1}{2}\gamma_p\xi_i(t)^2 \end{cases} \\ &\text{If } \lambda_i(t) > \frac{\beta}{1+\beta}\overline{\lambda}(t) \quad \Rightarrow \quad \begin{cases} z_i(t+1) = z_i(t) + \frac{1}{R}[\lambda_i(t) + \beta(\lambda_i(t) - \overline{\lambda}(t)) - \frac{1}{2}\gamma_p\hat{\beta}\lambda_i^2] \\ \lambda_i(t+1) = \lambda_i(t) - \xi_i(t) - \frac{1}{2}\gamma_p\xi_i(t)^2 \end{cases} \end{aligned} \tag{24}$$

where $\hat{\beta} = 1 + \beta + \beta^2$. Therefore, in this limit, the evolution of λ decouples from that of ζ (or z). Note that from Eq. 24 one has $-1 - \frac{\beta}{1+\beta}\overline{\lambda}(t) - \gamma_p/2 < \lambda < 1 - \frac{\beta}{1+\beta}\overline{\lambda}(t) - \gamma_p/2$ (which we rewrite in the form $\lambda_{\min} < \lambda < \lambda_{\max}$).

From the simplified evolution equation Eq. 24 we now obtain the evolution of the probability distribution $P_t(\lambda)$ and for $\overline{z}(t)$, which are by definition:

$$P_{t+1}(\lambda) = \int_0^1 d\xi \int_{\frac{\beta}{1+\beta}\overline{\lambda}(t)}^{\lambda_{\max}} d\lambda' P_t(\lambda')\delta\left(\lambda - \lambda' + \xi + \frac{\gamma_p}{2}\xi^2\right) + \int_0^1 d\xi \int_{\lambda_{\min}}^{\frac{\beta}{1+\beta}\overline{\lambda}(t)} d\lambda' P_t(\lambda')\delta\left(\lambda - \lambda' - \xi + \frac{\gamma_p}{2}\xi^2\right) \ , \quad (25)$$

and

$$\overline{z}(t+1) = \overline{z}(t) + \frac{1}{R}\int_{\frac{\beta}{1+\beta}\overline{\lambda}(t)}^{\lambda_{\max}} d\lambda\, P_t(\lambda)\, \left(\lambda + \beta(\lambda - \overline{\lambda}) - \frac{1}{2}\gamma_p\hat{\beta}\lambda^2\right) - \int_{\lambda_{\min}}^{\frac{\beta}{1+\beta}\overline{\lambda}(t)} d\lambda\, P_t(\lambda)\, \min\left\{-\lambda - \beta(\lambda - \overline{\lambda}) + \frac{1}{2}\gamma_p\hat{\beta}\lambda^2, \overline{z}(t)\right\} . \quad (26)$$

Since the dynamics of λ is decoupled from the one of $\overline{z}$, we can assume that $P_t(\lambda)$ reaches a stationary state. In Appendix C 1 we show that, at first order in γ_p, we have for the stationary distribution

$$P_{\rm st}(\lambda) = \left(1 - |\lambda + \frac{\beta\gamma_p}{4}| - \frac{\gamma_p}{2}\mathrm{sgn}(\lambda)\lambda^2\right) . \quad (27)$$

Hence, using the condition $P_{\rm st}(\lambda) \geq 0$ we find that $|\lambda| \leq 1 + O(\gamma_p)$, consistently with the assumptions we made above. From this result, we also find that $\overline{\lambda} = -\gamma_p(1+\beta)/4$.

Note that the average price here decreases as β increases, confirming the discussion in Section III C 2. Indeed, we have neglected the fluctuations of S in the above calculation, while, as explained in Section III C 2, it is the increase of involuntary savings (and thus of S) with β that leads to the average price increase. As a further confirmation, one can easily simulate Eqs. (24) with $\alpha_i \neq 0$, and find $d\overline{\lambda}/d\beta > 0$ in this case.

Let us now discuss the dynamics of $\overline{z}(t)$. If at some time t we have $\overline{z}(t) = 0$, then it is clear from Eq. (26) that $\overline{z}(t)$ will grow with time, and it will continue to do so unless the last term in the equation becomes sufficiently large. Recalling that $|\lambda| \leq 1 + O(\gamma_p)$, it is clear that even if $\overline{z}(t)$ becomes very large, the last term in Eq. (26) can be at most given by $\int_{-\infty}^{0} d\lambda\, P_{\rm st}(\lambda)\, \left(-\lambda - \beta(\lambda - \overline{\lambda}) + \frac{1}{2}\gamma_p\lambda^2\right)$. Therefore, if

$$\frac{1}{R}\int_{-\gamma_p\beta/4}^{\lambda_{\max}} d\lambda\, P_{\rm st}(\lambda)\, \left(\lambda + \beta(\lambda - \overline{\lambda}) - \frac{1}{2}\gamma_p\hat{\beta}\lambda^2\right) > \int_{\lambda_{\min}}^{-\gamma_p\beta/4} d\lambda\, P_{\rm st}(\lambda)\, \left(-\lambda - \beta(\lambda - \overline{\lambda}) + \frac{1}{2}\gamma_p\hat{\beta}\lambda^2\right) , \quad (28)$$

then $\overline{z}(t)$ will continue to grow with time, and $\overline{Y}(t) = 1 - \gamma_p\eta_+\overline{z}(t)$ will become very small and the economy will collapse. Conversely, if the condition in Eq. (28) is not satisfied, then Eq. (26) admits a stationary solution with $0 < \overline{z}_{\rm st} < \infty$. Using Eq. (27), the critical boundary line finally reads:[18]

$$R_c = \frac{\eta_+}{\eta_-} = \frac{\int_{-\gamma_p\beta/4}^{\lambda_{\max}} d\lambda\, P_{\rm st}(\lambda)\, \left(\lambda + \beta(\lambda - \overline{\lambda}) - \frac{1}{2}\gamma_p\hat{\beta}\lambda^2\right)}{\int_{\lambda_{\min}}^{-\gamma_p\beta/4} d\lambda\, P_{\rm st}(\lambda)\, \left(-\lambda - \beta(\lambda - \overline{\lambda}) + \frac{1}{2}\gamma_p\hat{\beta}\lambda^2\right)} \approx 1 - \gamma_p\frac{(2+\beta)^2}{2(1+\beta)} + ... \quad (29)$$

where we used the condition $\gamma_p < 4/(\beta+2)$ which is always verified for $0 < \beta < +\infty$ and $\gamma_p \to 0$. As one can see in Fig. 3 this results is in good agreement with numerical result for $\beta = 2$. We therefore find, interestingly, that increasing β *decreases* the value of R_c, i.e. stabilizes the high employment phase, as indeed discussed above, see Fig. 4 and the discussion around it.

In other words, the ζ_i perform a *biased* random walk, in presence of a noise whose average is given by the last two terms in Eq. (26) (of course, when $\overline{\zeta}$ is too small the minimum in the last term is important, because it is there to prevent $\overline{\zeta}$ from becoming negative). The system will evolve towards full employment, with $\overline{\zeta} = O(\eta_+)$ and $\overline{Y} = 1 - O(\gamma_p\eta_+)$, whenever the average noise is negative, and to full collapse, with $\overline{\zeta} \to \infty$ and $\overline{Y} = 0$, otherwise. The critical line is given by the equality condition, such that the average noise vanishes. Therefore, right at the critical point, the unemployment rate makes an unbiased random walk in time, meaning that its temporal fluctuations are large and slow. This justifies the "adiabatic" approximation[19] made above, that lead us to neglect the dynamics of the savings.

---

[18] It would be interesting to compute the first non trivial corrections in η to the transition line. We leave this for a later study.

[19] This refers, in physics, to a situation where a system is driven by an infinitely slow process. One can then consider the system to be always close to equilibrium during the process.

B. Oscillations in the high employment phase

As discussed above, in the FE phase macroeconomic variables display an oscillatory dynamics, see Fig. 5. Intuitively, the mechanism behind these oscillations is the following. When prices are low, demand is higher than production and firms increase the prices. But at the same time, households cannot consume what they demand, so they involuntarily save: savings increase when prices are low. These savings keep the demand high for a few rounds even while prices are increasing, therefore prices keep increasing above their equilibrium value. When prices are too high, households need to use their savings to consume, and therefore savings start to fall. Increase of prices and decrease of savings determine a contraction of the demand. At some point demand falls below production and prices start to decrease again, with savings decreasing at the same time. When prices are low enough, demand becomes again higher then production and the cycle is restarted. An example is shown in Fig. 5.

Based on this argument, it is clear that to study these oscillations, we need to take into account the dynamics of the savings so we cannot assume $\alpha_i = 0$ as in the previous section. However, here it is enough to consider the first order terms in γ_p. In terms of the basic variables in Eq. (20), the other variables that appear in Eq. (19) are easily written as follows:

$$\begin{aligned} \overline{Y}(t) &= 1 - \gamma_p \overline{\zeta}(t) \ , \\ s(t) &= 1 + \gamma_p \overline{\alpha}(t) \ , \\ \overline{p}(t) &= 1 + \gamma_p \overline{\lambda}(t) \ , \end{aligned} \tag{30}$$

where for ζ, α, λ, overlines denote flat averages over firms, and

$$D_i(t) = 1 + \gamma_p \left(\frac{1}{2}\overline{\alpha}(t) - \frac{1}{2}\overline{\zeta}(t) - \lambda_i(t) \right) \ . \tag{31}$$

Inserting this in Eq. (19), we arrive to the following equations that hold at the lowest order in γ_p:

$$\begin{aligned} &\text{If } -\zeta_i(t) < \frac{1}{2}\overline{\alpha}(t) - \frac{1}{2}\overline{\zeta}(t) - \lambda_i(t) \quad \Rightarrow \quad \begin{cases} \zeta_i(t+1) = \zeta_i(t) - \min\{\eta_+[\frac{1}{2}\overline{\alpha}(t) - \frac{1}{2}\overline{\zeta}(t) - \lambda_i(t) + \zeta_i(t)], \overline{\zeta}(t)\} \\ \text{If } \lambda_i(t) < \overline{\lambda}(t) \quad \Rightarrow \quad \lambda_i(t+1) = \lambda_i(t) + \xi_i(t) \\ \alpha_i(t+1) = \alpha_i(t) - \lambda_i(t) \end{cases} \\ &\text{If } -\zeta_i(t) > \frac{1}{2}\overline{\alpha}(t) - \frac{1}{2}\overline{\zeta}(t) - \lambda_i(t) \quad \Rightarrow \quad \begin{cases} \zeta_i(t+1) = \zeta_i(t) - \eta_-[\frac{1}{2}\overline{\alpha}(t) - \frac{1}{2}\overline{\zeta}(t) - \lambda_i(t) + \zeta_i(t)] \\ \text{If } \lambda_i(t) > \overline{\lambda}(t) \quad \Rightarrow \quad \lambda_i(t+1) = \lambda_i(t) - \xi_i(t) \\ \alpha_i(t+1) = \alpha_i(t) - \zeta_i(t) - \frac{1}{2}\overline{\alpha}(t) + \frac{1}{2}\overline{\zeta}(t) \end{cases} \end{aligned} \tag{32}$$

From an analytical point of view, the above model is still to complex to make progress. We make therefore a further simplification, by assuming that

$$\zeta_i = C\lambda_i \ , \tag{33}$$

where C is a certain numerical constant. In terms of the original Mark 0 variables, this approximation is equivalent to, roughly speaking,

$$p_i - \overline{p} \propto (Y_i - D_i) - (\overline{Y} - \overline{D}). \tag{34}$$

i.e. fluctuations of the prices are proportional to supply-demand gaps. Although numerical simulations only show a weak correlation, the approximation (33) allows us to obtain a more tractable model that retains the basic phenomenology of the oscillatory cycles and reads:

$$\begin{aligned} &\text{If } 2(1-C)\lambda_i(t) < \overline{\alpha}(t) - C\overline{\lambda}(t) \quad \Rightarrow \quad \begin{cases} \text{If } \lambda_i(t) < \overline{\lambda}(t) \quad \Rightarrow \quad \lambda_i(t+1) = \lambda_i(t) + \xi_i(t) \\ \alpha_i(t+1) = \alpha_i(t) - \lambda_i(t) \end{cases} \\ &\text{If } 2(1-C)\lambda_i(t) > \overline{\alpha}(t) - C\overline{\lambda}(t) \quad \Rightarrow \quad \begin{cases} \text{If } \lambda_i(t) > \overline{\lambda}(t) \quad \Rightarrow \quad \lambda_i(t+1) = \lambda_i(t) - \xi_i(t) \\ \alpha_i(t+1) = \alpha_i(t) - C\lambda_i(t) - \frac{1}{2}\overline{\alpha}(t) + \frac{1}{2}C\overline{\lambda}(t) \end{cases} \end{aligned} \tag{35}$$

This very minimal model, when simulated numerically, indeed gives persistent oscillations, independent on N, when $C > C^* \approx 0.45$, and can also be partially investigated analytically, see Appendix C 1.

C. The "representative firm" approximation

To conclude this section, we observe that there is a further simplification that allows one to retain some of the phenomenology of Mark 0. It consists in describing the firm sector by a unique "representative firm", $N_{\rm F} = 1$, with production $\overline{Y}(t)$, price $p(t)$ and demand $\overline{Y}(t)/p(t)$. The dynamics of the production and price are given by the same rule as above, but now the dynamics of the price completely decouples:

$$\begin{aligned} p(t) < 1 \;\Rightarrow\; & \begin{cases} \overline{Y}(t+1) = \overline{Y}(t)(1+\eta_+(\frac{1}{p(t)}-1)) \\ p(t+1) = p(t)(1+\gamma\xi(t)) \end{cases} \\ p(t) > 1 \;\Rightarrow\; & \begin{cases} \overline{Y}(t+1) = \overline{Y}(t)(1+\eta_-(\frac{1}{p(t)}-1)) \\ p(t+1) = p(t)(1-\gamma\xi(t)). \end{cases} \end{aligned} \tag{36}$$

Of course, this simple model misses several important effects: most notably those associated to Θ, hence the transition between the Full Employment, Endogenous Crises, and Residual Unemployment phases in Fig. 4. In particular, endogenous crises are never present in this case, because of the absence of a bankrupt/revival mechanism, and also the oscillatory pattern in the Full Employment phase disappears, because in this model savings are not considered. Still, this model is able to capture the transition between the Full Unemployment and Full Employment regions as a function of R (see Fig. 4), as confirmed by the analytical solution, which is in fact identical to the one of the model with $N_{\rm F} > 1$ when $\beta = 0$ and η is small. And since the model is so simple, one can hope that some of the extensions discussed in the previous section can be at least partly understood analytically within this "representative firm" framework (we will give a few explicit. This would be an important step to put the rich phenomenology that we observe on a firmer basis.

VI. SUMMARY, CONCLUSION

The aim of our work (which is part of the CRISIS project and still ongoing) was to explore the possible types of *phenomena* that simple macroeconomic Agent-Based Models can reproduce, and to map out the corresponding *phase diagram* of these models, as Figs. 7 and 4 exemplify. The precise motivation for our study was to understand in detail the nature of the macro-economic fluctuations observed in the "Mark I" model devised by D. Delli Gatti and collaborators [25, 26]. One of our central findings is the generic existence, in Mark I (and variations around that model) of a first order, discontinuous phase transition between a "good economy" where unemployment is low, and a "bad economy" where unemployment is high. By studying a simpler hybrid model (Mark 0), where the household sector is described by aggregate variables and not at the level of agents[20], we have argued that this transition is induced by an *asymmetry between the rate of hiring and the rate of firing* of the firms. This asymmetry can have many causes at the micro-level, for example different hiring and firing costs. In Mark I, for example, it reflects the reluctance of firms to take loans when the interest rate is too high. As the interest rate increases, the unemployment level remains small until a tipping point beyond which the economy suddenly collapses. If the parameters are such that the system is close to this transition, any small fluctuations (for example in the level of interest rates) is amplified as the system jumps between the two equilibria. It is actually possible that the central bank policy (absent in our current model), when attempting to stabilize the economy, in fact bring the system close to this transition. Indeed, too low an interest rate leads to overheating and inflation, and too high an interest rate leads to large unemployment. The task of the central bank is therefore to control the system in the vicinity of an instability and could therefore be a natural realization of the enticing 'self-organized criticality' scenario recently proposed in [46] (see also [19]).

Mark 0 is simple enough to be partly amenable to analytic treatments, that allow us to compute approximately the location of the transition line as a function of the hiring/firing propensity of firms, and characterize the oscillations and the crises that are observed. Mark 0 can furthermore be extended in several natural directions. One is to allow this hiring/firing propensity to depend on the financial fragility of firms – hiring more when firms are financially healthy and firing more when they are close to bankruptcy. We find that in this case, the above transition survives but becomes *second order*. As the transition is approached, unemployment fluctuations become larger and larger, and the corresponding correlation time becomes infinite, leading to very low frequency fluctuations. There again, we are able to give some analytical arguments to locate the transition line [43]. Other stabilizing mechanisms, such as

[20] For a recent study exploring the idea of hybrid models, see [45].

the bankruptcy of indebted firms and their replacement by healthy firms (financed by the accumulated savings of households), lead to a similar phenomenology.

The role of the bankruptcy threshold Θ, which is the only "proto-monetary" effect in Mark 0, turns out to be crucial in the model, and leads to the phase diagram shown in Fig. 4. We generically find not one but three "good" phases of the economy: one is where full employment prevails (FE), the second one is where some residual unemployment is present (RU), and the third, intermediate one is prone to *acute endogenous crises* (EC), during which the unemployment rate shoots up before the economy recovers. Note again that there are no exogeneous productivity shocks in Mark 0; interestingly, crises (when they occur) are of endogeneous origin and not as a result of sweeping parameters through a phase transition (as is the case, we argued, of Mark I with a time dependent interest rate) but purely as a result of the own dynamics of the system. The existence of endogenous crises driven by feedback loops in such simple settings is quite interesting from a general standpoint (see also [29, 31]), and reinforces the idea that many economic and financial crises may *not* require large exogenous shocks (see [5, 6, 26, 47, 50–54] for related discussions). We have shown that the endogenous crises can be defanged (in the model) if the household sector does not carry the full burden of firms bankruptcies and/or when the profits of firms is efficiently re-injected in the economy (see e.g. Fig. 6).

We have then allowed firms to vary the wages of their employees according to some plausible rules of thumb (wages in Mark I and Mark 0 are fixed); this leads to inflation or deflation but leaves the above picture essentially unchanged (see Fig. 7). Several other extensions are of obvious interest, and we plan to study them in the near future in the same stylized way as above. The most obvious ones is to understand how a central bank that prints money and sets exogenously the interest rate can control the unemployment rate and the inflation rate in the vicinity of an unstable point, as we mentioned just above [43]. Other interesting topics are: modeling research and innovation, allowing firms to produce different types of goods, and introducing a financial sector and a housing market [55].

Beyond the generic phase diagram discussed in the whole paper, we found another notable, robust feature: the low unemployment phase of all the ABMs we considered are characterized by endogenous oscillations that do not vanish as the system size becomes large, with a period corresponding, in real time, to $\sim 5-10$ years [31]. It is tempting to interpret these oscillations as real and corresponding to the "business cycle", as they arise from a very plausible loop between prices, demand and savings. These oscillations actually also appear in highly simplified models, where both the household and the firm sectors are represented by aggregate variables [37, 56], as well as in network models [6].

Building upon this last remark, a very important question, it seems to us, is how much can be understood of the phenomenology of ABMs using "mean-field" approaches [58, 59], i.e. dynamical equations for aggregate variables of the type considered, for example, in [37, 38]? A preliminary analysis reveals that the dynamical equations corresponding to Mark 0 or Mark I already lead to an amazingly complex phase-diagram [56]. Are these mean-field descriptions quantitatively accurate? When do we really need agents and when is an aggregate description sufficient? The answer to this question is quite important, since it would allow one to devise faithful "hybrid" ABMs, where whole sectors of the economy would be effectively described in terms of these aggregate variables, only keeping agents where they are most needed.

Another nagging question concerns the calibration of macroeconomic ABMs. It seems to us that before attempting any kind of quantitative calibration, exploring and making a catalogue of the different possible qualitative "phases" of the model is mandatory. Is the model qualitatively plausible or is the dynamics clearly unrealistic? In what "phase" is the true economy likely to be? On this point, one of the surprise of the present study is the appearance of very long time scales. For example, even in the case of perfectly stable economy with wage update rule (18) and all γ parameters equal to 10% (a rather large value), the equilibrium state of the economy (starting from an arbitrary initial condition) is only reached after ≈ 200 time steps. If one thinks that the elementary time scale in these models is of the order of three months, this means that the physical equilibration time of the economy is 20-50 years, or even much longer, see e.g. Figs. 4 & 7. But there is no reason to believe that on these long periods all the micro-rules and their associated parameters are stable in time. Therefore, studying the *stationary state* of macroeconomic ABMs might be completely irrelevant to understand the real economy. The economy could be in a perpetual transient state (aka "non ergodic"), unless one is able to endogenise the time evolution of all the relevant parameters governing its evolution (see the conclusion of [13] for a related discussion).

If this is the case, is there any use in studying ABMs at all? We strongly believe that ABMs would still be genuinely helpful. ABMs allow us to experiment and scan the universe of possible outcomes – not missing important scenarios is already very good macroeconomics. Human imagination turns out to be very limited, and that is the reason we like models and equations, that help us guessing what *can* happen, especially in the presence of collective effects that are often very counterintuitive. In this respect, ABMs provide extremely valuable tools for generating scenarios, that can be used to test the effect of policy decisions (see e.g. the pleas by Buchanan [21], and Farmer & Foley [22]). In order to become more quantitative, we think that ABMs will have to be calibrated at the micro-level, using detailed behavioural experiments and field studies to fine tune all the micro-rules needed to build the economy from bottom up (see [55, 57] for work in this direction.) Calibrating on historical data the emergent macro-dynamics of ABMs

will most probably fail, because of the dimensionality curse and of the Lucas critique (i.e. the feedback between the trajectory of the economy and policy decisions that dynamically change the parameters of the model).

Acknowledgements

This work was partially financed by the CRISIS project. We want to thank all the members of CRISIS for most useful discussions, in particular during the CRISIS meetings. The input and comments of P. Aliferis, T. Assenza, J. Batista, E. Beinhocker, M. Buchanan, D. Challet, D. Delli Gatti, D. Farmer, J. Grazzini, C. Hommes, F. Lillo, G. Tedeschi, S. Battiston, A. Kirman, A. Mandel, M. Marsili and A. Roventini are warmly acknowledged. We are especially grateful to F. Altarelli and G. Cencetti for a careful critical reading of the manuscript and for their suggestions. The comments and criticisms of the referees were also useful to improve the manuscript. Finally, J.P.B. wants to thank J.C. Trichet for his warm support for this type of research when he was president of the ECB.

Appendix A: Pseudo-code for Mark I+

We describe here the pseudo-code of our version of Mark I, which we call Mark I+. To keep the length reasonable, a few irrelevant details will be omitted, but the information given here is enough to reproduce the results presented in the paper. In particular, we describe here only the part of the code that is needed to generate the dynamical evolution of the model, and we omit the part that is needed to generate the data output. The source code is available on the site of the CRISIS project (www.crisis-economics.eu).

1. Notations

We describe the algorithm in an object-oriented fashion, where the different agents are described as representatives of a few classes. We use an object oriented syntax. This syntax should be very intuitive and easy to follow. However, it is useful to clarify a few notational conventions:

- The declaration of a variable a (for example of integer type `int`) will be written in C syntax as `int` a.
- If a is an object of some class, then $a.f(x)$ means that we are calling the method (function) $f(x)$ of object a with argument x.
- For simplicity, in the **for** loops we will use the C syntax where, for example, **for**$(t \leftarrow 1; t \leq T; t \leftarrow t+1)$ means that t is set to one before the loop starts, t is increased by 1 at the end of each iteration, and the loop continues if the condition $t \leq T$ is true.
- Instead of arrays we will use *vectors* of objects, and we will follow the notation of C++[21]. For example, `vector<int>` will denote an ordered set (array) of integers. Moreover:
  - A vector of size N will be declared as `vector<int>` $A(N)$, and by default the declaration `vector<int>` A means that the set A is initiated as empty.
  - $|A|$ will denote the size of the set A
  - The notation $A[i]$ will denote the i-th element of the set (with $i = 0, \cdots, |A|-1$).
  - The notation $A \leftarrow a$ will denote the operation of adding the element a to the set A, therefore increasing its size by one (this correspond to the "push_back" operation in C++).
  - The notation $a \leftarrow_R A$ will denote the extraction of a random element from A, which is set equal to a. Note that the element is not removed from A so the size of A remains constant. The notation $A' \leftarrow_{R,M} A$ denotes extraction at random of M different elements from the set A, that constitute the set A'.
  - The notation $A \not\leftarrow a$ will denote the removal of element a from set A, therefore reducing the size of A by one.
  - To denote an ordered iteration over an ordered set A, we will use a loop **for**$(a \in A)$.
  - We define a function average$(a.f(),\ a \in A)$ that returns the average of the $|A|$ numbers $a[i].f()$, $i = 0, \cdots, |A|-1$
- For logical operators, we use the C++ convention[22] which should be quite transparent.

2. Classes

The classes are:

- *Firm*
- *Household*
- *Bank*

---

[21] See for example `http://www.cplusplus.com/reference/vector/vector/`
[22] `http://en.wikipedia.org/wiki/Operators_in_C_and_C++#Logical_operators`

The role of the bank is very limited at this level and this class is mainly included for the purpose of future extension. One household has a special role as it is the "owner" of the firms, which will pay dividends to this one household. The main loop is described in Algorithm 1 below. The implementation of the firm class is in Algorithm 3, the household class in Algorithm 6 and the bank class in Algorithm 7.

Algorithm 1 Main loop of Mark I+

Require: N_{F} Number of firms; N_{H} Number of households; ρ_0 baseline interest rate; T total evolution time;

▷ **Initialisation**

`vector<household>` $H(N_{\mathrm{H}})$
`household` O ▷ The owner of all firms
`vector<firm>` $F(N_{\mathrm{F}})$
`bank` B

$\overline{p} \leftarrow 1$ ▷ Average price
for $f \in F$ **do**
  f.set_owner(O)
end for
B.set_ρ(ρ_0)

▷ **Main loop**

for $(t \leftarrow 1; t \leq T; t \leftarrow t+1)$ **do**

  `vector<firm>` E,D
  for $f \in F$ **do** ▷ **Firm decide new strategy on prices and production**
    f.set_new_strategy($\overline{p}$)
    f.get_loans(B)
    f.compute_interests()
    f.define_labor_demand()
    if f.n_vacancies() > 0 **then** $D \leftarrow f$ ▷ Firms in D demand workforce
    else if f.n_vacancies() < 0 **then** $E \leftarrow f$ ▷ Firms in E have an excess of workforce
    end if
  end for

  for $f \in E$ **do** ▷ **Job market and production**
    while f.n_vacancies() < 0 **do** f.fire_random_worker() ▷ Firms with excess workforce fire random workers
    end while
  end for
  `vector<household>` U
  for $h \in H$ **do**
    if !h.working() **then** $U \leftarrow h$ ▷ U is the set of unemployed households
    end if
  end for
  while $|U| > 0$ && $|D| > 0$ **do** ▷ Random match of unemployed households and demanding firms
    $h \leftarrow_R U$
    $f \leftarrow_R D$
    f.hire(h)
    $U \not\rightarrow h$
    if f.n_vacancies()==0 **then** $D \not\rightarrow f$
    end if
  end while

Algorithm 2 Main loop of Mark I+ (continued)

for $f \in F$ **do** ▷ Firms produce and pay workers
  f.produce()
  f.pay_workers()
  if f.age()<100 **then** f.markup_rule() ▷ Young firms apply a markup rule to avoid bankrupt
  end if
end for

▷ **Goods market**

$H \leftarrow$random_permutation(H)
for $h \in H$ **do** h.consume(F) ▷ Consume in random order
end for

▷ **Accounting and bankrupts**

bad_debts $\leftarrow 0$
`vector<firm>` L
for $f \in F$ **do**
  f.accounting(B)
  if f.liquidity() < 0 **then** ▷ Firms with negative liquidity go bankrupt
    bad_debts $\leftarrow$ bad_debts + f.liquidity() ▷ Note: bad_debts is negative!
    $\overline{L} \leftarrow f$ ▷ $\overline{L}$ is the sent of bankrupt firms
  else $L \leftarrow f$ ▷ L is the set of healthy firms
  end if
end for
if $|L| == 0$ **then** `break` ▷ If all firms are bankrupt, exit the program
end if
$\overline{p}_b \leftarrow$ average($f[i]$.price(), $f \in L$)
$\overline{Y}^T \leftarrow$ average($f[i]$.target_production(), $f \in L$)
$\overline{Y} \leftarrow$ average ($f[i]$.production(), $f \in L$)
for $f \in \overline{L}$ **do** f.reinit($\overline{p}_b, \overline{Y}^T, \overline{Y}$) ▷ Bankrupt firms are reinitialized with the average parameters of healthy firms
end for
total_liquidity $\leftarrow \sum_{i=0}^{N_F} f[i]$.equity() $+ \sum_{i=0}^{N_H} h[i]$.wealth()
for $f \in F$ **do** f.get_money(bad_debts * $f[i]$.equity() / total_liquidity)
end for
for $h \in H$ **do** h.get_money(bad_debts * $h[i]$.wealth() / total_liquidity)
end for ▷ Bad debt is spread over firms and households proportionally to their wealth
$\overline{p} \leftarrow \frac{\sum_{i=0}^{|F|-1} f[i].\text{price}()\, f[i].\text{sales}()}{\sum_{i=0}^{|F|-1} f[i].\text{sales}()}$ ▷ Update average price
end for

Algorithm 3 The class `firm`

Parameters : $W = 1$, $\alpha = 1$, $\gamma_p = 0.1$, $\gamma_y = 0.1$, $\mu = 0$, $\delta = 0.2$, $\tau = 0.05$
Dynamic variables : `vector<household>` E; `household` $O; p, Y, Y^T, D, \mathcal{L}, v, \mathcal{D}^T, t$
Dynamic variables (auxiliary): $L_d, \rho, \mathcal{I}$

▷ **Initialization methods**

function INIT
  $E \leftarrow$empty ▷ The set E is the list of employees and is initialized as empty
  $p \leftarrow 1$ ▷ Price
  $Y \leftarrow 1$ ▷ Production
  $Y^T \leftarrow 1$ ▷ Target production
  $D \leftarrow 1$ ▷ Demand
  $\mathcal{L} \leftarrow 50$ ▷ Liquidity
  $v \leftarrow 0$ ▷ Number of vacancies
  $\mathcal{D}^T \leftarrow 0$ ▷ Total debt
  $t \leftarrow 0$ ▷ Internal clock
end function

function SET_OWNER(`household` $\widetilde{O}$)
  $O \leftarrow \widetilde{O}$
end function

function REINIT($\widetilde{p}, \widetilde{Y}^T, \widetilde{Y}$)
  $p \leftarrow \widetilde{p}$
  $Y \leftarrow \widetilde{Y}$
  $Y^T \leftarrow \widetilde{Y}^T$
  $D \leftarrow 0$
  $\mathcal{L} \leftarrow \min\{O.\text{wealth}(), Y/\alpha\}$ ▷ The owner injects money to restart the bankrupt firm
  O.get_money(-$\mathcal{L}$)
  $v \leftarrow 0$
  $\mathcal{D}^T \leftarrow 0$
  $t \leftarrow 0$
  $v \leftarrow 0$
  for $h \in E$ **do**
    fire(h)
  end for
end function

▷ **Output methods**

function PRICE
  return p
end function

function PRODUCTION
  return Y
end function

function STOCK
  return $Y - D$
end function

function SALES
  return D
end function

function TARGET_PRODUCTION
  return Y^T
end function

function EQUITY
  return $\mathcal{L} - \mathcal{D}^T$
end function

function LIQUIDITY
  return $\mathcal{L}$
end function

function N_VACANCIES
  return v
end function

Algorithm 4 The class `firm` (continued)

function AGE
  return t
end function

▷ **Accounting methods**

function GET_MONEY($\widetilde{m}$)
  $\mathcal{L} \leftarrow \mathcal{L} + \widetilde{m}$
end function

function GET_LOANS(`bank` $\widetilde{B}$)
  $\mathcal{L}_n \leftarrow WL_d - \mathcal{L}$ ▷ Financial need
  if $\mathcal{L}_n > 0$ **then**
    $\ell \leftarrow (\mathcal{D}^T + \mathcal{L}_n)/(\mathcal{L} + 0.001)$ ▷ This is the leverage
    $\rho_{\text{offer}} \leftarrow \widetilde{B}$.compute_offer_rate($\ell$) ▷ New offered interest rate
    $\mathcal{D}^c \leftarrow \mathcal{L}_n\, F(\rho_{\text{offer}})$ ▷ The function $F(\rho)$ can be whatever decreasing function of ρ, see Eq. (4)
    if $\mathcal{D}^c > 0$ **then**
      $\rho \leftarrow \rho_{\text{offer}}$ ▷ If new credit is contracted, the interest rate is updated
      $\mathcal{D}^T \leftarrow \mathcal{D}^T + \mathcal{D}^c$ ▷ Total debt is increased by current debt $\mathcal{D}^c$
      $\mathcal{L} \leftarrow \mathcal{L} + \mathcal{D}^c$
      $\widetilde{B}$.get_money(-$\mathcal{D}^c$)
    end if
  end if
end function

function COMPUTE_INTERESTS
  $\mathcal{I} \leftarrow \rho \mathcal{D}^T$ ▷ Interests to be paid in this round
end function

function PAY_WORKERS
  for $h \in E$ **do**
    h.get_money(W)
  end for
  $\mathcal{L} \leftarrow \mathcal{L} - W|E|$
end function

function MARKUP_RULE
  if $Y > 0$ **then**
    $p_{\text{markup}} \leftarrow (1+\mu)(W|E| + \mathcal{I})/Y$
    $p \leftarrow \max\{p, p_{\text{markup}}\}$
  end if
end function

function ACCOUNTING(`bank` $\widetilde{B}$)
  $\mathcal{L} \leftarrow \mathcal{L} - \mathcal{I} - \tau \mathcal{D}^T$ ▷ Firm pays interests and repays a fraction τ of its debt
  $\widetilde{B}$.get_money($\mathcal{I} + \tau \mathcal{D}^T$)
  $\mathcal{D}^T \leftarrow (1-\tau)\mathcal{D}^T$
  $\mathcal{P} \leftarrow pD - W|E| - \mathcal{I}$ ▷ Profit
  if $\mathcal{P} > 0$ **then**
    O.get_money($\delta \mathcal{P}$) ▷ Firm pays dividends to the owner
    $\mathcal{L} \leftarrow \mathcal{L} - \delta \mathcal{P}$
  end if
end function

function SELL($\widetilde{q}$)
  $D \leftarrow D + \widetilde{q}$
  $\mathcal{L} \leftarrow \mathcal{L} + p\widetilde{q}$
end function

Algorithm 5 The class `firm` (continued)

▷ **Production and job market methods**

function SET_NEW_STRATEGY($\widetilde{p}$)
  $t \leftarrow t + 1$
  if $Y = D$ && $p \geq \widetilde{p}$ **then** $Y^T \leftarrow Y(1 + \gamma_y \, \texttt{random})$ ▷ This is Eq. (1) in the main text
  else if $Y = D$ && $p < \widetilde{p}$ **then** $p \leftarrow p(1 + \gamma_p \, \texttt{random})$
  else if $Y > D$ && $p \geq \widetilde{p}$ **then** $p \leftarrow p(1 - \gamma_p \, \texttt{random})$
  else if $Y > D$ && $p < \widetilde{p}$ **then** $Y^T \leftarrow Y(1 - \gamma_y \, \texttt{random})$
  end if
  $Y^T \leftarrow \max\{Y^T, \alpha\}$
  $L_d \leftarrow \text{ceil}(Y^T/\alpha)$
end function

function DEFINE_LABOR_DEMAND
  $L_d \leftarrow \min\{L_d, \text{floor}(\mathcal{L}/W)\}$
  $L_d \leftarrow \max\{L_d, 0\}$
  $v \leftarrow L_d - |E|$
end function

function PRODUCE
  $Y = \min\{Y^T, \alpha|E|\}$
  $D = 0$ ▷ The demand is reset to zero at each production cycle
end function

function HIRE(`household` h)
  $E \leftarrow h$
  h.get_job(W)
  $v \leftarrow v - 1$
end function

function FIRE(`household` h)
  h.lose_job()
  $E \not\rightarrow h$
  $v \leftarrow v + 1$
end function

function FIRE_RANDOM_WORKER
  if $|E| > 0$ **then**
    $h \leftarrow_R E$
    fire(h)
  end if
end function

Algorithm 6 The class `household`

Parameters : $M = 3$, $c = 0.8$
Dynamic variables : S, W

▷ **Initialization methods**

function INIT
  $S \leftarrow 0$ ▷ Savings
  $W \leftarrow 0$ ▷ Salary
end function

▷ **Accounting methods**

function GET_MONEY($\widetilde{m}$)
  $S \leftarrow S + \widetilde{m}$
end function

▷ **Output methods**

function WEALTH
  return S
end function

function WORKING
  if $W > 0$ **then return** True
  else return False
  end if
end function

▷ **Job and goods market methods**

function GET_JOB($\widetilde{W}$)
  $W \leftarrow \widetilde{W}$
end function

function LOSE_JOB
  $W \leftarrow 0$
end function

function CONSUME(`vector<firm>` $\widetilde{F}$)
  budget $\leftarrow cS$
  if budget > 0 **then**
    $F_c \leftarrow_{R,M} \widetilde{F}$ ▷ Extract M random firms from $\widetilde{F}$ and put them in the set F_c
    $F_c \leftarrow$ sort($f \in F_c$, f.price()) ▷ Order the set F_c according to firms' prices
    spent $\leftarrow 0$
    for ($i \leftarrow 0; i < |F_c|$ && spent $<$ budget$; i \leftarrow i + 1$) **do**
      $s \leftarrow f[i]$.stock() ▷ s is the stock available from this firm
      if $s > 0$ **then**
        $q \leftarrow$ (budget - spent)$/f[i]$.price() ▷ Maximum possible consumption from this firm
        if $s > q$ **then**
          $f[i]$.sell(q)
          spent $\leftarrow$ budget ▷ The household has finished the budget, the loop ends
        else
          $f[i]$.sell(s)
          spent $\leftarrow$ spent $+ s\, f[i]$.price()
        end if
      end if
    end for
  end if
  $S \leftarrow S-$spent
end function

Algorithm 7 The class `bank`

Dynamic variables : E, ρ_b

function INIT
  $E \leftarrow 0$ ▷ Bank liquidity
  $\rho_b \leftarrow 0$ ▷ Baseline interest rate
end function

function SET_$\rho(\widetilde{\rho})$
  $\rho_b \leftarrow \widetilde{\rho}$
end function

function COMPUTE_OFFER_RATE$(\widetilde{\ell})$
  return $\rho_b\, G(\widetilde{\ell})$ ▷ We chose $G(\ell) = 1 + \log(1 + \ell)$
end function

function GET_MONEY$(\widetilde{m})$
  $E \leftarrow E + \widetilde{m}$
end function

Appendix B: Pseudo-code of Mark 0

We present here the pseudo-code for the Mark 0 model discussed in Sec. III A and in Sec. IV. The source code is available on the site of the CRISIS project (www.crisis-economics.eu).

Algorithm 8 The basic Mark 0

Require: $N_{\rm F}$ Number of firms; $\mu, c, \beta, \gamma_p, \eta_+^0, \eta_-^0, \delta, \Theta, \varphi, f$; T total evolution time;

▷ **Initialization**

for $(i \leftarrow 0; i < N_{\rm F}; i \leftarrow i+1)$ **do**
  $W[i] \leftarrow 1$ ▷ Salaries are always fixed to one
  $p[i] \leftarrow 1 + 0.2(\texttt{random} - 0.5)$
  $Y[i] \leftarrow \mu[1 + 0.2(\texttt{random} - 0.5)]/2$ ▷ Initial employment is 0.5
  $\mathcal{E}[i] \leftarrow W[i]Y[i]\,2\,\texttt{random}$
  $a[i] \leftarrow 1$
end for

$S \leftarrow N_{\rm F} - \sum_i \mathcal{E}[i]$

▷ **Main loop**

for $(t \leftarrow 1; t \leq T; t \leftarrow t+1)$ **do**
  $u \leftarrow 1 - \frac{1}{\mu N_{\rm F}} \sum_i Y[i]$
  $\varepsilon \leftarrow 1 - u$
  $\overline{p} \leftarrow \frac{\sum_i p[i]Y[i]}{\sum_i Y[i]}$
  $\overline{w} \leftarrow \frac{\sum_i W[i]Y[i]}{\sum_i Y[i]}$
  $\tilde{u}[i] \leftarrow \frac{\exp(\beta W[i]/\overline{w})}{\sum_i a[i]\exp(\beta W[i]/\overline{w})} N_{\rm F} u$
  ▷ **Firms update prices, productions and wages**
  for $(i \leftarrow 0; i < N_{\rm F}; i \leftarrow i+1)$ **do**

    if $a[i] == 1$ **then**

      if $Y[i] < D[i]$ **then** ▷ Wage update
        if $\mathcal{P}[i] > 0$ **then**
          $W[i] \leftarrow W[i][1 + \gamma_w \varepsilon\ \texttt{random}]$
          $W[i] \leftarrow \min\{W[i], P[i]\min[D[i], Y[i]]/Y[i]\}$
        end if
        $Y[i] \leftarrow Y[i] + \min\{\eta_+(D[i] - Y[i]), \mu\tilde{u}[i]\}$ ▷ This is Eq. (11) in the main text
        if $p[i] < \overline{p}$ **then** $p[i] \leftarrow p[i](1 + \gamma_p\,\texttt{random})$
        end if

      else if $Y[i] > D[i]$ **then**
        if $\mathcal{P}[i] < 0$ **then**
          $W[i] \leftarrow W[i][1 - \gamma_w u\ \texttt{random}]$
        end if
        $Y[i] \leftarrow \max\{0, Y[i] - \eta_-(D[i] - Y[i])\}$
        if $p[i] < \overline{p}$ **then** $p[i] \leftarrow p[i](1 - \gamma_p\,\texttt{random})$
        end if
      end if
    end if
  end for

  $u \leftarrow 1 - \frac{1}{\mu N_{\rm F}} \sum_i Y[i]$ ▷ Update u and $\overline{p}$
  $\overline{p} \leftarrow \frac{\sum_i p[i]Y[i]}{\sum_i Y[i]}$

Algorithm 9 The basic Mark0 (continued)

▷ **Households decide the demand**

$C \leftarrow c(\max\{S, 0\} + \sum_i W[i]Y[i])$

for $(i \leftarrow 0; i < N_{\rm F}; i \leftarrow i + 1)$ **do**
  $D[i] \leftarrow \frac{Ca[i]\exp(-\beta p[i]/\overline{p})}{p[i]\sum_i a[i]\exp(-\beta p[i]/\overline{p})}$ ▷ Inactive firms have no demand
end for

▷ **Accounting**

for $(i \leftarrow 0; i < N_{\rm F}; i \leftarrow i + 1)$ **do**
  if $a[i] == 1$ **then**
    $\mathcal{P}[i] \leftarrow p[i]\min\{Y[i], D[i]\} - W[i]Y[i]$
    $S \leftarrow S - \mathcal{P}[i]$
    $\mathcal{E}[i] \leftarrow \mathcal{E}[i] + \mathcal{P}[i]$
    if $\mathcal{P}[i] > 0$ && $\mathcal{E}[i] > 0$ **then** ▷ Pay dividends
      $S \leftarrow S + \delta\,\mathcal{P}[i]$
      $\mathcal{E}[i] \leftarrow \mathcal{E}[i] - \delta\,\mathcal{P}[i]$
    end if
    if $\mathcal{E}[i] > \Theta W[i]Y[i]$ **then** ▷ Set of healthy firms
      $\mathcal{H} \leftarrow i$
    end if
  end if
end for

▷ **Defaults**

deficit $= 0$
for $(i \leftarrow 0; i < N_{\rm F}; i \leftarrow i + 1)$ **do**
  if $a[i] == 1$ && $\mathcal{E}[i] < -\Theta Y[i]W[i]$ **then**
    $j \leftarrow_R \mathcal{H}$
    if `random` $< 1 - f$ && $\mathcal{E}[j] > -\mathcal{E}[i]$ **then** ▷ Bailed out
      $\mathcal{E}[j] \leftarrow \mathcal{E}[j] + \mathcal{E}[i]$
      $\mathcal{E}[i] \leftarrow 0$
      $p[i] \leftarrow p[j]$
      $W[i] \leftarrow W[j]$
    else ▷ Bankrupted
      deficit $\leftarrow$ deficit $-\mathcal{E}[i]$
      $a[i] \leftarrow 0$
      $Y[i] \leftarrow 0$
      $\mathcal{E}[i] \leftarrow 0$
    end if
  end if
end for

▷ **Revivals**

$\mathcal{E}^+ \leftarrow 0$
for $(i \leftarrow 0; i < N_{\rm F}; i \leftarrow i + 1)$ **do**
  if $a[i] == 0$ && `random` $< \varphi$ **then** ▷ Reactivate firm
    $a[i] \leftarrow 1$
    $p[i] \leftarrow \overline{p}$
    $Y[i] \leftarrow \mu u$ `random`
    $\mathcal{E}[i] \leftarrow W[i]Y[i]$
    deficit $\leftarrow$ deficit $+\mathcal{E}[i]$
  end if
  if $a[i] == 1$ && $\mathcal{E}[i] > 0$ **then** ▷ Firms total savings
    $\mathcal{E}^+ \leftarrow \mathcal{E}^+ + \mathcal{E}[i]$
  end if
end for

▷ **Debt**

if deficit $> S$ **then** ▷ Households cannot be indebted
  deficit $\leftarrow$ deficit $-S$
  $S \leftarrow 0$
  for $(i \leftarrow 0; i < N_{\rm F}; i \leftarrow i + 1)$ **do**
    if $a[i] == 1$ && $\mathcal{E}[i] > 0$ **then**
      $\mathcal{E}[i] \leftarrow \mathcal{E}[i] - \mathcal{E}[i]/\mathcal{E}^+$ deficit
    end if
  end for
else
  $S \leftarrow S -$ deficit
end if

end for

Appendix C: Perturbative solution of the schematic model

1. Stationary distribution at first order in γ_p

From Eq. (25), the stationary distribution $P_{\rm st}(\lambda)$ satisfies

$$P_{\rm st}(\lambda) = \int_0^1 d\xi \int_0^\infty d\lambda' P_{\rm st}(\lambda')\delta\left(\lambda - \lambda' + \xi + \frac{\gamma_p}{2}\xi^2\right) + \int_0^1 d\xi \int_{-\infty}^0 d\lambda' P_{\rm st}(\lambda')\delta\left(\lambda - \lambda' - \xi + \frac{\gamma_p}{2}\xi^2\right) . \tag{C1}$$

We recall that the dynamics of $\lambda_i(t)$ is

$$\begin{cases} \lambda_i(t) < 0 & \lambda_i(t+1) = \lambda_i(t) + \xi_i(t) - \frac{1}{2}\gamma_p \xi_i(t)^2 \\ \lambda_i(t) > 0 & \lambda_i(t+1) = \lambda_i(t) - \xi_i(t) - \frac{1}{2}\gamma_p \xi_i(t)^2 \end{cases} \tag{C2}$$

with $\xi_i(t)$ a random variable in $[0,1]$. From these equations it is clear that positive and negative λ_i are pushed towards the origin, and after some transient time one necessarily has

$$-1 - \frac{1}{2}\gamma_p \le \lambda_i(t) \le 1 - \frac{1}{2}\gamma_p . \tag{C3}$$

We first set $\gamma_p = 0$ and therefore restrict $\lambda \in [-1,1]$ in Eq. (C1). We obtain

$$P_{\rm st}(\lambda) = \int_{\max(0,-\lambda)}^{1-\max(\lambda,0)} {\rm d}\xi P_{\rm st}(\lambda+\xi) + \int_{\max(0,\lambda)}^{1-\max(-\lambda,0)} {\rm d}\xi P_{\rm st}(\lambda-\xi) = \mathcal{L}_0 P_{\rm st} , \tag{C4}$$

where we call $\mathcal{L}_0$ the linear operator that appears on the right hand side. When one computes the action of $\mathcal{L}_0$ on the functions $g_n(\lambda) = {\rm sign}(\lambda)\lambda^n$ and $h_n(\lambda) = \lambda^n$, one finds, for $\lambda > 0$:

$$(n+1)\mathcal{L}_0 g_n(\lambda) = \begin{cases} 1 - (-1)^n - \sum_{k=0}^{n-1} D_n^k (-1)^k \lambda^{n-k} & \lambda > 0 \\ 1 - (-1)^n + \sum_{k=0}^{n-1} D_n^k \lambda^{n-k} & \lambda < 0 \end{cases} \tag{C5}$$

with C_n^k the binomial coefficients and $D_n^k = \sum_{j=k}^n C_j^k$, and

$$(n+1)\mathcal{L}_0 h_n(\lambda) = \begin{cases} 1 - \lambda^{n+1} - (\lambda-1)^{n+1} = 1 + (-1)^n - 2\lambda^{n+1} - \sum_{k=1}^n C_{n+1}^k (-1)^k \lambda^{n+1-k} & \lambda > 0 , \\ (1+\lambda)^{n+1} + \lambda^{n+1} + (-1)^n = 1 + (-1)^n + 2\lambda^{n+1} + \sum_{k=1}^n C_{n+1}^k \lambda^{n+1-k} & \lambda < 0 . \end{cases} \tag{C6}$$

Let us focus on small values of n, useful in the following:

$$\begin{aligned} n=0 \;&\to\; \mathcal{L}_0 g_0 = 0 , \qquad && \mathcal{L}_0 h_0 = 2(h_0 - g_1) , \\ n=1 \;&\to\; \mathcal{L}_0 g_1 = 1 - g_1 , \qquad && \mathcal{L}_0 h_1 = -g_2 + h_1 , \\ n=2 \;&\to\; \mathcal{L}_0 g_2 = h_1 - g_2 , \qquad && \mathcal{L}_0 h_2 = \frac{2}{3}(h_0 - g_3) + h_2 - g_1 . \end{aligned} \tag{C7}$$

In particular, one has:

$$\mathcal{L}_0(h_0 - g_1) = 2(h_0 - g_1) - 1 + g_1 = (h_0 - g_1). \tag{C8}$$

This shows that $P_0(\lambda) = h_0(\lambda) - g_1(\lambda) = 1 - |\lambda|$ (called the "tent") is an eigenvector of $\mathcal{L}_0$ with eigenvalue 1, i.e. this is the stationary state for $\gamma_p = 0$.

This basic solution allows us to obtain perturbatevely the stationary solution for small γ_p. Recalling that the support of $P_{\rm st}(\lambda)$ is given by Eq. (C3), Eq. (C1) can be written as

$$P_{\rm st}(\lambda) = \int_{a_-}^{a_+} {\rm d}\xi P_{\rm st}(\lambda + \xi + \frac{1}{2}\gamma_p\xi^2) + \int_{b_-}^{b_+} {\rm d}\xi P_{\rm st}(\lambda - \xi + \frac{1}{2}\gamma_p\xi^2) = \mathcal{L}_{\gamma_p} P_{\rm st} , \tag{C9}$$

with:

$$\begin{aligned} a_- &= \max\left\{0, -\lambda(1 + \frac{\gamma_p}{2}\lambda)\right\} , \quad a_+ = \min\left\{1, 1 - \lambda - \frac{\gamma_p}{2}(1 + (1-\lambda)^2)\right\} , \\ b_- &= \max\left\{0, \lambda(1 + \frac{\gamma_p}{2}\lambda)\right\} , \quad b_+ = \min\left\{1, 1 + \lambda + \frac{\gamma_p}{2}(1 + (1+\lambda)^2)\right\} , \end{aligned} \tag{C10}$$

these integration bounds coming from the combination of the support in Eq. (C3) and the integration bounds on λ' in Eq. (C1). Writing $P_{\rm st} = P_0 + \gamma_p P_1$ where P_0 is the tent solution, and $\mathcal{L}_{\gamma_p} = \mathcal{L}_0 + \gamma_p \mathcal{L}_1$, we obtain at first order in γ_p an equation of the form:

$$(1 - \mathcal{L}_0) P_1 = \mathcal{L}_1 P_0 = S \ . \tag{C11}$$

The source term S can be most easily computed by computing $\mathcal{L}_{\gamma_p} P_0$ and expanding the result at first order in γ_p. We find $S = \lambda/2 - \mathrm{sign}(\lambda)\lambda^2 = \frac{1}{2} h_1 - g_2$. This is nice because one can look for a solution involving only $n = 1$ and $n = 2$ that closes the equation:

$$P_1 = \alpha h_1 + \beta g_2 \tag{C12}$$

Using: $\mathcal{L}_0 g_2 = h_1 - g_2 = \mathcal{L}_0 h_1$, one finds:

$$(1 - \mathcal{L}_0) P_1 = \alpha h_1 + \beta g_2 - (\alpha + \beta)(h_1 - g_2) = -\beta h_1 + (\alpha + 2\beta) g_2 = S = \frac{1}{2} h_1 - g_2 \ , \tag{C13}$$

so the result for the coefficients is

$$\beta = -\frac{1}{2} \ , \qquad \alpha = 0 \ . \tag{C14}$$

The final solution for the stationary state to first order in γ is:

$$P = 1 - |\lambda| - \frac{1}{2} \gamma_p \, \mathrm{sign}(\lambda) \lambda^2 . \tag{C15}$$

Note that P is still normalized, as it should and goes to zero (to first order in γ_p) at the boundary of the support interval given in Eq. (C3).

2. Perturbative analysis of the oscillations

In order to understand oscillations, we start from Eqs. (35) and in the following we assume that $0 < C < 1$. We introduce $\mathcal{E}(t) = \frac{\overline{\alpha}(t) - C\overline{\lambda}(t)}{2(1-C)}$, $\Lambda(t) = \min\{\overline{\lambda}(t), \mathcal{E}(t)\}$ and $\Omega(t) = \max\{\overline{\lambda}(t), \mathcal{E}(t)\}$, and we write Eq. (35) equivalently as

$$\begin{aligned} &\alpha_i(t+1) = \alpha_i(t) - \min\{\lambda_i(t), C\lambda_i(t) + \frac{1}{2}(\overline{\alpha}(t) - C\overline{\lambda}(t))\} \\ &\begin{cases} \text{If } \lambda_i(t) < \Lambda(t) & \Rightarrow \quad \lambda_i(t+1) = \lambda_i(t) + \xi_i(t) \\ \text{If } \Omega(t) < \lambda_i(t) < \Lambda(t) & \Rightarrow \quad \lambda_i(t+1) = \lambda_i(t) \\ \text{If } \lambda_i(t) > \Omega(t) & \Rightarrow \quad \lambda_i(t+1) = \lambda_i(t) - \xi_i(t) \end{cases} \end{aligned} \tag{C16}$$

The master equation for the distribution of λ reads:

$$\begin{aligned} P_{t+1}(\lambda') &= \int_{-\infty}^{\Lambda(t)} d\lambda \int_0^1 d\xi \, P_t(\lambda) \, \delta(\lambda' - \lambda - \xi) + \int_{\Lambda(t)}^{\Omega(t)} d\lambda \, P_t(\lambda) \delta(\lambda - \lambda') + \int_{\Omega(t)}^{\infty} d\lambda \int_0^1 d\xi \, P_t(\lambda) \, \delta(\lambda' - \lambda + \xi) \\ &= \int_{-\infty}^{\Lambda(t)} d\lambda \int_0^1 d\xi \, P_t(\lambda) \, \delta(\lambda' - \lambda - \xi) + \theta(\Lambda(t) \le \lambda' \le \Omega(t)) P_t(\lambda') + \int_{\Omega(t)}^{\infty} d\lambda \int_0^1 d\xi \, P_t(\lambda) \, \delta(\lambda' - \lambda + \xi) \ . \end{aligned} \tag{C17}$$

while the evolution equation for $\overline{\alpha}$ is

$$\overline{\alpha}(t+1) = \overline{\alpha}(t) - \int_{-\infty}^{\infty} d\lambda \, P_t(\lambda) \min\{\lambda, C\lambda + \frac{1}{2}(\overline{\alpha}(t) - C\overline{\lambda}(t))\} \ . \tag{C18}$$

a. Stationary state

Numerically, we observe that $\Lambda(t) \sim \Omega(t)$ and their variations are much smaller than the width of the distributions of λ and α. In the limit where

$$\Lambda(t) = \Omega(t) = \overline{\lambda}(t) = \mathcal{E}(t) = \frac{\overline{\alpha}(t) - C\overline{\lambda}(t)}{2(1-C)} \tag{C19}$$

we obtain

$$P_{t+1}(\lambda') = \int_{-\infty}^{\Lambda(t)} d\lambda \int_0^1 d\xi \, P_t(\lambda)\, \delta(\lambda' - \lambda - \xi) + \int_{\Lambda(t)}^{\infty} d\lambda \int_0^1 d\xi \, P_t(\lambda)\, \delta(\lambda' - \lambda + \xi) \ , \tag{C20}$$

whose stationary solution is $P(\lambda) = P_0[\lambda - \overline{\lambda}^*]$ where $P_0(x) = (1-|x|)\theta(1-|x|)$ is the tent function, with $\overline{\lambda}^*$ undetermined at this stage. Plugging this result in Eq. (C18), we have

$$\overline{\alpha}(t+1) = \overline{\alpha}(t) - \overline{\lambda}^* + \frac{1-C}{6} \ . \tag{C21}$$

The fixed point is therefore $\overline{\lambda}^* = \frac{1-C}{6}$ and from the condition (C19) we get $\overline{\alpha}^* = (2-C)\overline{\lambda}^* = (2-C)(1-C)/6$.

b. Oscillation around the stationary state – a simple approximation

In order to study the small oscillations we can make a very simple approximation, namely that at each time t we have $P_t(\lambda) = P_0[\lambda - \bar{\lambda}(t)]$. If we inject this approximation in Eq. (C18) we get, provided $\bar{\lambda}(t) - \overline{\lambda}^*$ is not too large, that

$$\begin{aligned} \bar{\alpha}(t+1) &= \bar{\alpha}(t) - \int_{-\infty}^{\infty} d\lambda \min\{\lambda, C\lambda + (1-C)\mathcal{E}(t)\} P_0[\lambda - \bar{\lambda}(t)] \\ &= \bar{\alpha}(t) - \bar{\lambda}(t) - \frac{1-C}{6}\left[-1 + 3A - 3A^2 + A^3 \operatorname{sign}(A)\right]_{A=\mathcal{E}(t)-\bar{\lambda}(t)} \end{aligned} \tag{C22}$$

Next, injecting this approximation in Eq. (C17) we get

$$\begin{aligned} \bar{\lambda}(t+1) &= \bar{\lambda}(t) + \frac{1}{2}\int_{-\infty}^{\Lambda(t)} d\lambda \, P_0[\lambda - \bar{\lambda}(t)] - \frac{1}{2}\int_{\Omega(t)}^{\infty} d\lambda \, P_0[\lambda - \bar{\lambda}(t)] \\ &\int_{\Omega(t)-\bar{\lambda}(t)}^{\infty} dx \, T[x] \\ &= \frac{1}{2}\left[\Lambda(t) + \Omega(t) - \frac{1}{2}[\Lambda(t) - \bar{\lambda}(t)]^2 \mathrm{sgn}[\Lambda(t) - \bar{\lambda}(t)] - \frac{1}{2}[\Omega(t) - \bar{\lambda}(t)]^2 \mathrm{sgn}[\Omega(t) - \bar{\lambda}(t)]\right] \end{aligned} \tag{C23}$$

We therefore get a closed system of two equations for $\bar{\lambda}(t)$ and $\bar{\alpha}(t)$. For small C, this system of equations converges quickly to the fixed point. However, for C sufficiently close to 1 ($C \sim 0.94$) it has a limit cycle of period 2, followed by an exponential divergence.

c. Oscillation around the stationary state – perturbative computation

Another strategy to characterize the oscillations is to look again for a perturbed solution around the stationary state, with: $\overline{\lambda}(t) = \frac{1-C}{6} + \overline{\delta\lambda}_1(t)$ and $\overline{\alpha}(t) = (2-C)\overline{\lambda} + 2(1-C)\overline{\delta\lambda}_2(t) + C\overline{\delta\lambda}_1(t)$. The distribution of λ_i at time t is now $P_t(\lambda) = P_0(\lambda - \frac{1-C}{6}) + \delta P_t(\lambda - \frac{1-C}{6})$. The shifted variable $\lambda - \frac{1-C}{6}$ will be denoted x. The aim is to write an evolution for δP_{t+1} after one time step. The $\mathcal{L}_0$ operator is the same as before, as well as the set of functions g_n and h_n. We also introduce $\Delta(x)$ as a δ-function slightly spread out, but of unit area (its precise width is irrelevant if it is small enough).

To lowest order, one has:

$$\mathcal{L}_0 \Delta \approx \frac{1}{2} h_0. \tag{C24}$$

We will also need, as found above:

$$\mathcal{L}_0 h_0 = 2(h_0 - g_1), \qquad \mathcal{L}_0 g_0 = 0, \qquad \mathcal{L}_0 g_1 = h_0 - g_1. \tag{C25}$$

We now introduce $d = |\overline{\delta\lambda}_2 - \overline{\delta\lambda}_1|$ and $s = \overline{\delta\lambda}_2 + \overline{\delta\lambda}_1$. Assuming $d, s \ll 1$, one finds, to first order:

$$\delta P_{t+1} = \mathcal{L}_0 \delta P_t - \frac{1}{2} d_t h_0 + \frac{1}{2} s_t g_0 + d_t \Delta \tag{C26}$$

Note that the integral $[\delta P] = \int_{-1}^{1} dx \delta P(x)$ is zero and conserved in time, as it should be (one has $[h_0] = 2$, $[g_0] = 0$ and $[\Delta] = 1$).

The idea now is to expand δP in terms of the h, g functions. It turns out that only three of them are needed, plus the Δ contribution. Indeed, assume:

$$\delta P_t = A_t(h_0 - 2\Delta) + B_t g_0 + C_t(g_1 - \frac{1}{2} h_0) \tag{C27}$$

Then, using the dynamical equation and the algebra above, one finds:

$$\mathcal{L}_0 \delta P_t = A_t(2(h_0 - g_1) - h_0) = -2A_t(g_1 - \frac{1}{2} h_0) \tag{C28}$$

Hence, the evolution is closed on itself, with the following evolution rules:

$$A_{t+1} = -\frac{1}{2} d_t, \qquad B_{t+1} = \frac{1}{2} s_t, \qquad C_{t+1} = -2A_t. \tag{C29}$$

Note that $A_t - B_t = -\frac{1}{2}(d_{t-1} + s_{t-1}) = -\max(\overline{\delta\lambda}_{1t-1}, \overline{\delta\lambda}_{2t-1}) = -M_{t-1}$.

Now, by definition $\overline{\lambda}_t = \frac{1-C}{6} + \int_{-1}^{1} dx x \delta P_t$, but since h_0 and g_1 are even, the only contribution comes from the g_0 component. Therefore:

$$\overline{\delta\lambda}_{1t} = B_t \int_{-1}^{1} dx x g_0(x) = B_t. \tag{C30}$$

This leads to a first evolution equation:

$$\overline{\delta\lambda}_{1t+1} = \frac{1}{2} s_t = \frac{1}{2}(\overline{\delta\lambda}_{1t} + \overline{\delta\lambda}_{2t}). \tag{C31}$$

The second equation comes from the evolution of the α_i's. From:

$$\overline{\alpha}(t+1) = \overline{\alpha}(t) - C\overline{\delta\lambda}_1(t) - \frac{1-C}{2}\overline{\delta\lambda}_2(t) - (1-C)\int_{-1}^{0} dx\, x \delta P_t(x), \tag{C32}$$

we finally find:

$$\overline{\delta\lambda}_{2t+1} = \frac{3}{4}\overline{\delta\lambda}_{2t} - \frac{1}{4} M_{t-1} + \frac{7}{24} d_{t-2} - \frac{C}{4(1-C)}(\overline{\delta\lambda}_{1t} + \overline{\delta\lambda}_{2t}). \tag{C33}$$

The coupled set of iterations for $\overline{\delta\lambda}_{1t}$ and $\overline{\delta\lambda}_{2t}$, Eqs. (C31) and (C33), lead to damped oscillations for $C < C^{**}$ and sustained oscillations for $C > C^{**}$, with $C^{**} \approx 0.91$. The numerical value of C^{**} however does not coincide with that of $C^* \approx 0.45$, obtained from the direct simulation of Eqs. (35). This might be due to neglecting higher order corrections, which appear to be numerically large: the oscillations generated by Eqs. (35) are of amplitude ~ 0.1 at the onset, suggesting a sub-critical bifurcation. For small C, on the other hand, the oscillation amplitude is small and the above equations appear to be quantitatively correct, validating the above calculations.

---

[1] see e.g. S. Balibar, *The Discovery of Superfluidity*, Journal of Low Temperature Physics 146:441-470 (2007).

[2] A. Kirman, *Whom or What Does the Representative Individual Represent?*, Journal of Economic Perspectives, 6: 117-136 (1992).

[3] W. Brock, S. Durlauf, *Discrete choice with social interactions*, Rev. Econ. Stud. 68, 235 (2001)

[4] see e.g.: P. Ball. *Why Society is a Complex Matter*. Springer, Berlin (2012) for an introductory review.

[5] for a recent review, see: J.-P. Bouchaud, *Crises and Collective Socio-Economic Phenomena: Simple Models and Challenges*, J. Stat. Phys. 151: 567-606 (2013), and refs. therein.

[6] J. Bonart, J.-P. Bouchaud, A. Landier, D. Thesmar, *Instabilities in large economies: aggregate volatility without idiosyncratic shocks*, arXiv 1406.5022

[7] for a concise overview of ABM for collective phenomena, see: R. Goldstone, M. Janssen, *Computational Models of Collective Behavior*, Trends in Cognitive Science 9: 424-30 (2005).

[8] for early papers on ABM for economics, see e.g.: R. L. Axtell, *Why agents? On the varied motivations for agent computing in the social sciences*, in: Sallach, C. M. Macal and D. (Eds.), Proceedings of the Workshop on Agent Simulation, 2000; L. Tesfatsion, *Agent-Based Computational Economics: Growing Economies From Botton Up*, Artificial Life, 8: 55-82 (2002). More recent attempts can be found in: Q. Ashraf, B. Gershman and P. Howitt, *Banks, market organization, and macroeconomic performance: an agent-based computational analysis*. No. w17102. National Bureau of Economic Research (2011); M. Raberto, A. Teglio and S. Cincotti, *'Debt, deleveraging and business cycles: an agent-based perspective'* Economics: The Open-Access, Open-Assessment E-Journal 6.2012-27 (2012): 1-49; G. Dosi *et al.*, *'Income distribution, credit and fiscal policies in an agent-based Keynesian model'* Journal of Economic Dynamics and Control 37.8 (2013): 1598-1625.

[9] for some review papers on ABM for economics and finance, see e.g.: A. Leijonhufvud, *Agent-Based Macro* in Tesfatsion,L. and Judd, K. (eds.), Handbook of Computational Economics, 1625-1636 (2006); C. H. Hommes, *Heterogeneous Agent Models in Economics and Finance*, Handbooks in Economics Series 23, 1109-1186 North-Holland: Elsevier (2006); B. LeBaron and L. Tesfatsion, *Modeling macroeconomies as open-ended systems of interacting agents*, American Economic Review: Papers and Proceedings, 98: 246-250 (2008); H. Dawid and M. Neugart, *Agent-based Models for Economic Policy Design* Eastern Economic Journal, 37: 44-50 (2011). M. Cristelli, L. Pietronero, A. Zaccaria: *Critical overview of agent-based models for economics*. arXiv:1101.1847.

[10] C. Sims, *Macroeconomics and Reality*, Econometrica, 48: 1-48 (1980)

[11] see e.g. J. Gali, *Monetary Policy, Inflation and the Business Cycle*, Princeton University Press (2008).

[12] W. Buiter, *The unfortunate uselessness of most "state of the art" academic monetary economics*, http://blogs.ft.com/maverecon/2009/03/

[13] G. Fagiolo and A. Roventini, *'Macroeconomic policy in dsge and agent-based models'* Revue de l'OFCE 5 (2012): 67-116.

[14] G. Fagiolo, A. Moneta and P. Windrum, *A critical guide to empirical validation of agent-based models in economics: methodologies, procedures, and open problems*. Computational Economics 30.3 (2007): 195-226.

[15] A. Pyka and G. Fagiolo, *Agent-based modelling: a methodology for neo-Schumpeterian economics*. Elgar companion to neo-schumpeterian economics 467 (2007).

[16] A. Kirman, *The Economic Crisis is a Crisis for Economic Theory*, CESifo Economic Studies, 56: 498-535 (2010)

[17] R. J. Caballero, *Macroeconomics after the Crisis: Time to Deal with the Pretense-of-Knowledge Syndrome*, Journal of Economic Perspectives, 24: 85-102 (2010)

[18] for a general, insightful introduction, see: N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group*. AddisonWesley, Reading (1992)

[19] P. Bak, *How Nature Works: The Science of Self-Organized Criticality*, Copernicus, Springer, New York, 1996.

[20] on this point, see the interesting discussion in: I. Kondor, I. Csabai, G. Papp, E. Mones, G. Czimbalmos, M. Cs. Sandor, *Strong random correlations in networks of heterogeneous agents*, arXiv:1210.2224

[21] M. Buchanan, *This Economy does not compute*, New York Times, October 1, 2008.

[22] D. Farmer, D. Foley, *The economy needs agent-based modelling*, Nature 460, 685-686 (2009)

[23] I. Giardina, J.-P. Bouchaud, *Bubbles, Crashes and Intermittency in Agent Based Market Models*, The European Physics Journal B 31: 421-37 (2003).

[24] R. Lyea, J. Peng, L. Tana, S. A. Cheonga *Understanding agent-based models of financial markets: a bottom-up approach based on order parameters and phase diagrams*, arXiv:1202.0606;

[25] The Mark I family of models was elaborated in a series of papers and books, in particular: E. Gaffeo, D. Delli Gatti, S. Desiderio, and M. Gallegati, *Adaptive Microfoundations for Emergent Macroeconomics*, Eastern Economic Journal, 34: 441-463 (2008); D. Delli Gatti, E. Gaffeo, M. Gallegati, G. Giulioni, and A. Palestrini, *Emergent Macroeconomics: An Agent-Based Approach to Business Fluctuations*, Springer: Berlin, 2008.

[26] D. Delli Gatti, S. Desiderio, E. Gaffeo, P. Cirillo, M. Gallegati, *Macroeconomics from the Bottom-up*, Springer: Berlin (2011).

[27] L.Russo, *The forgotten revolution: How Science Was Born in 300 BC and Why It Had to Be Reborn*, Springer, Berlin (2004); Italian-speaking readers might also consult the wonderful first chapter of L.Russo, *Flussi e Riflussi: indagine sull'origine di una teoria scientifica*, Feltrinelli (2003), unfortunately out of print but available in many public libraries.

[28] C. Deissenberg, S. van der Hoog, H. Dawid, *Eurace: A massively parallel agent-based model of the european economy*, Applied Mathematics and Computation, 204: 541-552 (2008);

[29] H. Dawid, S. Gemkow, P. Harting, S. van der Hoog, M. Neugarty; *The Eurace@Unibi Model: An Agent-Based Macroeconomic Model for Economic Policy Analysis*, Working paper, Bielefeld University. 2011.

[30] G. Dosi, G. Fagiolo, A. Roventini, *An evolutionary model of endogenous business cycles*. Computational Economics, 27: 3-24 (2005)

[31] G. Dosi, G. Fagiolo, and A. Roventini, *Schumpeter meeting Keynes: A policyfriendly model of endogenous growth and business cycles*, Journal of Economic Dynamics and Control, 34, 1748-1767 (2010); and [13].

[32] A. Mandel, S. Furst, W. Lass, F. Meissner, C. Jaeger; *Lagom generiC: an agent-based model of growing economies*, ECF

Working Paper 1 /2009.
[33] C. Metzig, M. Gordon, *Heterogeneous Enterprises in a Macroeconomic Agent-Based Model*, arXiv:1211.5575.
[34] The precise version of Mark I that we used as a reference is the one presented by J. Grazzini and D. Delli Gatti at the Crisis Workshop in Milan, November 2012. Its transcription is given by the pseudo-code in Appendix A.
[35] Bank of England, *Quarterly Bulletin*, Q1 Vol. 54 No. 1 (2014).
[36] see e.g. S. P. Anderson, A. De Palma, J. F. Thisse, *Discrete Choice Theory of Product Differentiation*. MIT Press, New York (1992), and [5] for a recent discussion and further references.
[37] R. M. Goodwin, *A Growth Cycle*, in Feinstein, C.H. (ed) Socialism, Capitalism and Economic Growth, Cambridge University Press (1967)
[38] P. Flaschel, *The Macrodynamics of Capitalism*, Springer, Berlin 2010.
[39] F. Caccioli, P. Vivo and M. Marsili, *Eroding Market Stability by Proliferation of Financial Instruments*, European Physical Journal B, 71: 467-479 (2009).
[40] G. B. Eggertsson and P. Krugman, *Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach*, The Quarterly Journal of Economics 127, 1469-1513 (2012).
[41] B. C. Greenwald and J. E. Stiglitz, *Financial market imperfections and business cycles*. The Quarterly Journal of Economics 108, 77-114 (1993).
[42] M. Napoletano *et al.*, *Wage Formation, Investment Behavior and Growth Regimes: An Agent-Based Analysis* Revue de l'OFCE 5, 235-261 (2012).
[43] S. Gualdi, M. Tarzia, F. Zamponi, J.-P. Bouchaud, *Policy experiments in simple macro-economic ABM*, in preparation (CRISIS project).
[44] K. Anand, P. Gai, M. Marsili, *The rise and fall of trust networks*, Progress in Artificial Economics. Lect. Notes Econ. Math. Syst. 645: 77 (2010)
[45] T. Assenza, D. Delli Gatti, *E Pluribus Unum: Macroeconomic Modelling for Multi-agent Economies*, UCSC Working paper (2013).
[46] F. Patzelt and K. Pawelzik, *Criticality of Adaptive Control Dynamics*, Phys. Rev. Lett. 107: 238103 (2011)
[47] H. Minsky, *Stabilizing an Unstable Economy*, McGraw-Hill, New York (2008)
[48] M. W. L. Elsby *et. al*, *The ins and outs of cyclical unemployment*, American Economic Journal: Macroeconomics 1.1 (2009): 84-110.
[49] R. Shimer, *Reassessing the ins and outs of unemployment*, Review of Economic Dynamics 15.2 (2012): 127-148.
[50] D. Challet, M. Marsili and Y. C. Zhang, *Minority Games*, Oxford University Press, 2005.
[51] D. Sornette, *Endogenous versus exogenous origins of crises*. In: Albeverio, S., Jentsch, V., Kantz, H. (eds.) Extreme Events in Nature and Society. Springer, Heidelberg (2005)
[52] K. Anand, A. Kirman, M. Marsili, *Epidemics of rules, rational negligence and market crashes* Eur. J. Financ. (2011)
[53] J. Lorenz, S. Battiston, F. Schweitzer, *Systemic risk in a unifying framework for cascading processes on networks* Eur. Phys. J. B 71, 441 (2009)
[54] J.-P. Bouchaud *The endogenous dynamics of markets: price impact, feedback loops and instabilities* In: Berd, A. (ed.) Lessons from the 2008 Crisis. Risk Books, Incisive Media, London (2011)
[55] see e.g. J. Geanakoplos, R. Axtell, D. J. Farmer, P. Howitt, B. Conlee, J. Goldstein, M. Hendrey, N. M. Palmer, C.-Y. Yang, *Getting at Systemic Risk via an Agent-Based Model of the Housing Market*, Cowles Foundation Discussion Papers 1852, Cowles Foundation for Research in Economics, Yale University, 2012.
[56] S. Gualdi, M. Tarzia, F. Zamponi, J.-P. Bouchaud, *On simple aggregate descriptions of ABMs (and the real economy)*, unpublished report.
[57] see e.g. C. Hommes, *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems*, Cambridge University Press (2013) for a review on behavioral experiments.
[58] C. Di Guilmi, M. Gallegati and S. Landini, *Economic dynamics with financial fragility and mean-field interaction: A model* Physica A: Statistical Mechanics and its Applications 387.15 (2008): 3852-3861.
[59] M. Aoki and H. Yoshikawa, *Reconstructing macroeconomics: a perspective from statistical physics and combinatorial stochastic processes* Cambridge University Press, 2011.

Drug Discovery Today • Volume xxx, Number xx • xxxx 2021 REVIEWS

Artificial intelligence-enhanced drug design and development: Toward a computational precision medicine

Philippe Moingeon *, **Mélaine Kuenemann, Mickaël Guedj**

Servier, Research and Development, 50 rue Carnot, 92284 Suresnes Cedex, France

Artificial Intelligence (AI) relies upon a convergence of technologies with further synergies with life science technologies to capture the value of massive multi-modal data in the form of predictive models supporting decision-making. AI and machine learning (ML) enhance drug design and development by improving our understanding of disease heterogeneity, identifying dysregulated molecular pathways and therapeutic targets, designing and optimizing drug candidates, as well as evaluating *in silico* clinical efficacy. By providing an unprecedented level of knowledge on both patient specificities and drug candidate properties, AI is fostering the emergence of a computational precision medicine allowing the design of therapies or preventive measures tailored to the singularities of individual patients in terms of their physiology, disease features, and exposure to environmental risks.

Keywords: Artificial Intelligence; Big data; Computational precision medicine; Disease model; Drug discovery & development; Machine learning

Introduction

Drug development is a complex process that currently requires, on average, 12 years of development and a US$2.6 billion investment per individual drug made available to the patient.[1] Stringent regulatory requisites to demonstrate drug efficacy and safety result in a high attrition rate because of negative results during their evaluation in costly clinical studies, with an estimated 6.2% of drugs selected in the discovery phase eventually made available to patients.[2,3] In this context, **AI**-based predictive modeling (see Glossary) is emerging as a revolutionary solution to improve both the efficacy and speed of drug design and development, most particularly by optimizing early on the choice of therapeutic targets as well as drug candidates.[3–5] AI can be defined as a convergence of technologies recapitulating four dimensions of human intelligence (i.e., sensing, thinking, acting, and learning). As such, AI allows the integration of massive amounts of multi-modal data, both structured and unstructured, to build up probabilistic and dynamic models of a problem.

As it applies to drug development, AI-driven predictive models can be generated by using specific sets of data to inform a series of decisions taken throughout drug discovery, development, and registration steps (Fig. 1). These steps include selecting the right therapeutic target, the optimal drug candidate, the appropriate dosing and administration regimens, as well as the appropriate patients to include in clinical studies.[3,5] By providing a means to capture the value of data related to diagnosis, patient characterization, drug candidate attributes, and prediction of individual responses to therapy, AI enables a more personalized approach, termed 'precision medicine', that is proposing treatments better tailored to individual patient specificities.[6,7]

On this basis, we discuss herein four main applications of AI to support drug design and development: (i) the generation of disease models based on molecular profiling data from patients to represent disease heterogeneity; (ii) the identification of dysregulated molecular pathways and of candidate therapeutic targets predicted to contribute to disease causality; (iii) the design, synthesis, and optimization of drug candidates interacting with

* *Corresponding author.* Moingeon, P. (philippe.moingeon@servier.com)

https://doi.org/10.1016/j.drudis.2021.09.006

Please cite this article in press as: P. Moingeon et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.09.006

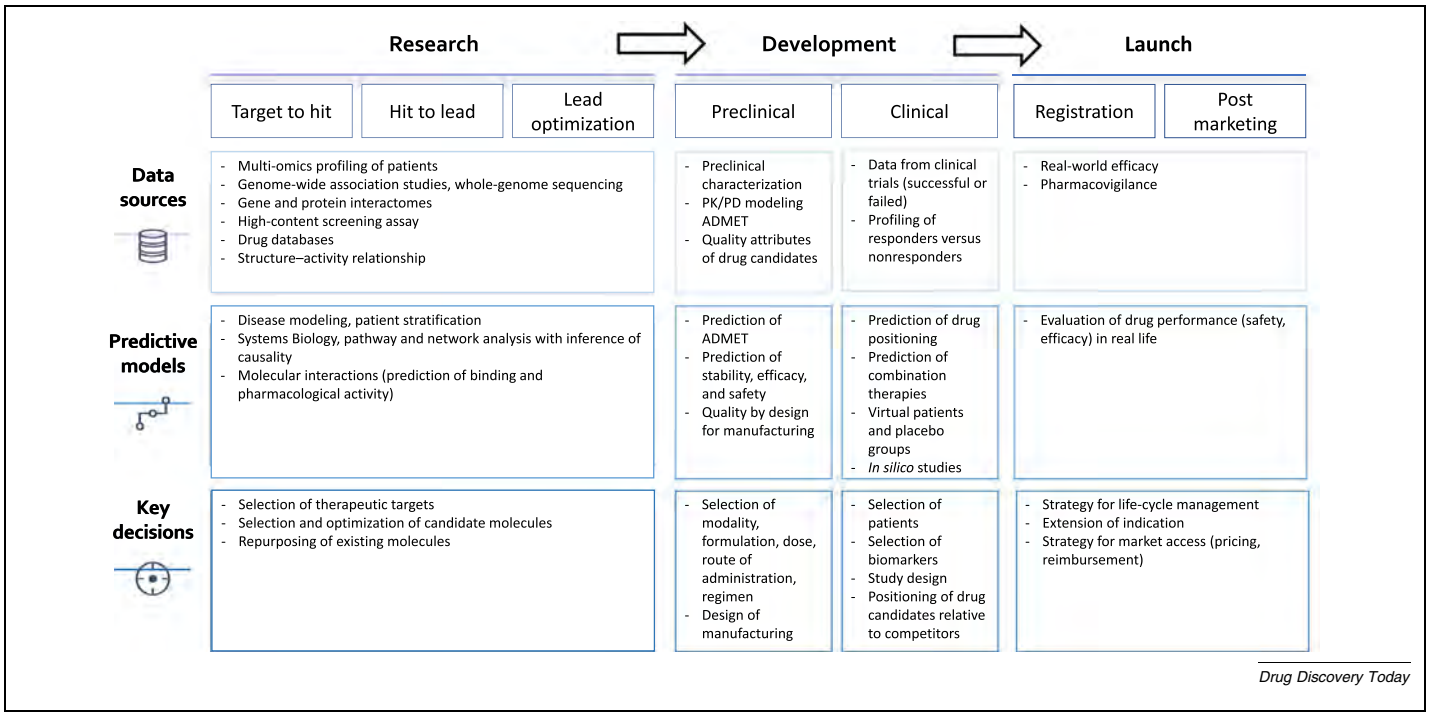

FIGURE 1

Decision-making during drug discovery, development, and registration. Key decisions to be taken as well as predictive models and examples of data sets supporting those models are provided for the drug discovery, development, and registration phases. Abbreviations: ADMET, absorption, distribution, metabolism, excretion, and toxicity; PD, pharmacodynamics; PK, pharmacokinetics.

these targets; and (iv) the evaluation of clinical efficacy by using virtual patients or real-world evidence data.

Capturing the value of big biomedical data

The recent and rapid advances in next-generation DNA, RNA, and exome sequencing, multi-omics molecular profiling, high-resolution medical imaging, and electronic capture technologies make it possible to characterize at an unprecedented level the specificities of individuals in terms of their physiology, pathophysiology of their disease as well as their environmental risk exposure. The Cancer Genome Atlas (TCGA), the Alzheimer's Disease Neuroimaging Initiative (ADNI), the Osteoarthritis Initiative (OAI), and the UK Biobank projects are all examples of this growing trend to integrate big data from large patient populations to support drug development (Table S1 in the supplemental information online). In the near future, such comprehensive molecular information will be available for millions of patients across multiple diseases, together with exponential data and knowledge compiled within hundreds of structured biomedical databases, such as those managed by the European Bioinformatics Institute (EBI) or the US National Center for Biotechnology Information (NCBI) (Table S1 in the supplemental information online).

When trying to capture the value of those ever-increasing amounts of data, the main challenges are linked to proper access and selection of standardized and machine-readable data, as well as to data complexity, heterogeneity, and **sparsity**. Integrating massive and multi-modal data generated from multiple technologies with proper quality attributes in terms of consistency and reliability remains a significant difficulty in data life-cycle management (Fig. 2). Access to accurate and curated data in large quantities is also key to improve **ML** repeatability. Solving those problems requires setting up computing hardware architectures adapted to life sciences specificities (Table S2 in the supplemental information online), often deported into the cloud. To this end, many initiatives, such as the Clinical Data Interchange Standards Consortium (CDISC)[8] or the FAIR guiding principles,[9] have emerged to enable the findability, accessibility, interoperability, reusability, and exchange of data.[10] In addition, the regulatory requirements imposed in terms of access, storage, sharing of confidential and sensitive health data by the European General Data Protection Regulation (GDPR)[11] and the US Health Information Technology for Economic and Clinical Act impose the implementation of clear and operational data governance strategies (Fig. 2).

In this context, precompetitive collaborative consortia between pharmaceutical companies or academic labs, such as MELLODDY[12] or the Drug Target Commons,[13] respectively, constitute innovative federated knowledge initiatives to assemble, curate, and share massive data of an appropriate quality for developing ML algorithms. The MELLODY consortium brings together several drug companies sharing their chemical libraries to train multitask predictive algorithms, subsequently applied by each individual partner in support of its own drug discovery program. In parallel, multiple crowd-source challenges, such as the Kaggle,[14] Dream,[15] and PrecisionFDA,[16] challenges propose data sets of reference to establish standards for benchmarking and testing novel algorithms to address complex biomedical problems.[6]

Please cite this article in press as: P. Moingeon et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.09.006

AI and disease modeling

The convergence of biotechnologies and AI provides an opportunity to create disease models to help positioning therapies in well-defined patient subpopulations. Such models are generated following extensive molecular profiling of patients compared with healthy controls using multi-omics technologies to represent diseases as endotypes defined based upon underlying pathophysiological mechanisms.[17] These data are classically produced during the follow-up of large cohorts of patients by public–private partnerships, with patient stratification being performed by using a combination of **unsupervised** and **supervised learning** approaches. The rationale for such a clustering, as a substitute to former classifications based solely upon clinical phenotypes, is that it better supports a precision medicine approach relying upon therapies targeted to well-defined subgroups of patients.[18] To this aim, molecular profiling data obtained in the blood and/or target organs of thousands of patients with a given disease are combined with detailed clinical information in terms of disease progression, severity, or response to treatments to stratify patients in homogeneous subgroups reflecting disease heterogeneity. Whereas the integration of such massive and multi-modal data is not possible with conventional bioinformatics, a comprehensive modeling of diseases can now be made by using AI.[19]

To do so, the main computational challenges still lie in the ability to: (i) integrate data coming from multi-omics technologies while reducing the multiplicity of their dimensions[20,21]; (ii) decipher disease mechanisms at a single cell level[22,23]; (iii) model the dynamic evolution of the disease[24]; and (iv) consolidate the findings through **consensus** and resampling approaches to support their validity and replication.[25] Following gene set enrichment analyses within each cluster, patient subgroups can then be further characterized in terms of molecular pathways being dysregulated.[26] Specific databases (e.g., Ingenuity Pathway Analysis and STRING) are used to regroup within established functional molecular pathways genes or proteins that are either up- or downregulated in patient samples compared with healthy controls. Given that a disease is defined in molecular terms in reference to normality, disease features need to be identified beyond molecular polymorphisms observed in association with the healthy state.

Overall, disease modeling can provide information on both the natural history of the disease and the relationship between pathophysiological mechanisms involved at both systemic and organ-specific levels. Furthermore, it sheds light on patient heterogeneity as well as on molecular signatures, which can be used to cluster patients in homogeneous groups to envision a precision medicine approach, taking into account patient specificities within clusters. Importantly, it also provides clues for further *in silico* identification of targets of therapeutic interest.

Identification, prioritization, and validation of therapeutic targets

Computational methods are being developed to identify disease-associated genes or proteins predicted to be involved in the causality of the disease, thus representing potential actionable therapeutic targets. As a first step, molecular pathways dysregulated in a given disease are represented in large-scale **networks** of interconnected genes or proteins, either established from protein–protein interactions (PPIs)[27] or reconstructed via inference techniques, such as correlation or Bayesian networks.[28,29] Such networks are also often referred to as **knowledge graphs**, representing knowledge as both concepts and relationships between them. This representation allows the delineation of disease-associated subnetwork modules serving as a basis for further computational analyses of their intrinsic topology to identify nodes predicted as 'causal' (including, for example, master regulators, hubs, and driver mutations).[19,30] In particular, **network propagation** algorithms (also referred to as diffusion) are commonly used to amplify the signal of nodes for which little or no direct evidence of disease association is available.[31] As above, main computational challenges involve the integration of multi**layer** networks obtained from different levels,[32] as well as the representation of large-scale dynamic information.[33]

Besides biological relevance, additional dimensions are considered to prioritize disease targets for further investigation, as

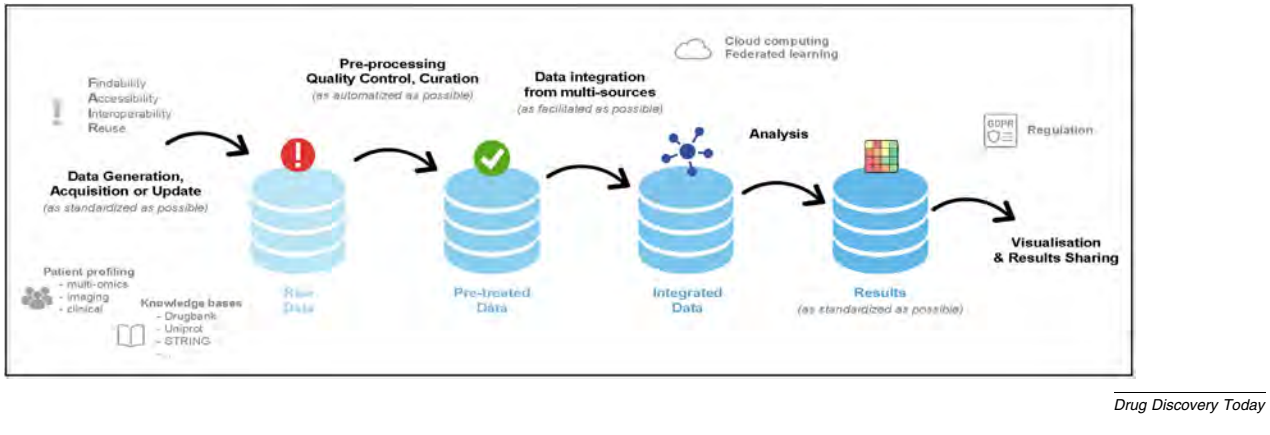

FIGURE 2

Biomedical data life-cycle management. Representation of a general biomedical data life cycle from data generation to the sharing of results, with emphasis on the needs for more standardization and automation in data governance.

Please cite this article in press as: P. Moingeon et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.09.006

illustrated by the Open Targets initiative.[34] This includes: (i) druggability (i.e., the likelihood of being able to modulate the function of a target with a small synthetic or biological drug, or any other therapeutic modality)[35]; (ii) potential safety implications when interfering with this target[36]; (iii) innovativeness documented from patent and literature mining by using **natural language processing** (NLP)[37]; and (iv) feasibility of drug development.[38] The selection of targets can be facilitated by the identification of features based on protein structure or sequence suggesting that they can bind small molecules.[4] Thus, the confirmation of target druggability significantly benefits from advances in 3D-structure modeling,[39] including the recent improved prediction of protein structures by DeepMind's AlphaFold algorithm based on primary amino acid sequences.[40]

Candidate targets identified with inferences of causality in the disease using network computing approaches need to be validated on the basis of empirical evidence generated in wet lab experiments. This validation step, which includes, for example, CRISPR-Cas9 gene deletion or siRNA gene silencing, phenotyping assessment of target expression on cells or tissues from patients relative to healthy controls or functional assays conducted in animal models, can be substantially simplified when using computational predictive models. As a result, both cost and timelines associated with drug discovery are reduced, while strengthening the rationale for selecting the candidate target before entering clinical development.

AI-enhanced drug design, selection, and optimization

Network-based proximity analyses allow the prediction of drug–target interactions, which can be applied to the repurposing of existing drugs in new indications.[41,42] For instance, the deepDTnet algorithm is based on a network-based **deep learning** methodology for *in silico* identification of new molecular targets for known drugs.[43] DeepDTnet embeds 15 types of chemical, genomic, phenotypic, and cellular networks to generate biologically and pharmacologically relevant features. AI is also generating considerable interest in the design or identification of new compounds with desirable properties from virtual drug target screens. Computational chemistry has been broadly used to document quantitative structure–activity relationships (QSAR) with the goal to predict activities in a chemical space potentially encompassing millions of molecules. The QSAR field benefited over the past decade from the combined application of deep learning to neural networks with higher computational power and better algorithms addressing the **overfitting** and **gradient problems**.[44,45] ML methods are now applied to train neural networks on ligand-based virtual screens to identify and optimize drugs interacting with candidate therapeutic targets, predict their absorption, distribution, metabolism, excretion, and toxicity (ADMET) characteristics, or repurpose existing molecules.[4,42,46]

Interestingly, deep learning allows multitask prediction by developing models encompassing more than one activity, such as bioactivity and ADME properties. Whereas the prediction of multiple activities can be trained in parallel, because they share the same input and hidden layers, each activity is associated with a specific output node (Fig. 3a). A Kaggle contest evaluating various ML approaches to improve the prediction performance of QSAR methods was won by a multitask deep network yielding a 15% improvement over baseline.[47] Besides improving the accuracy of the prediction, multitask prediction based on deep learning further enhances drug discovery compared with classical ML methods (such as Random Forrest or Support Vector Machine) because the latter only predict a single property at a time. Instead of solely relying upon on-the-shelf and expert-derived chemical features, deep learning also allows the identification of novel molecular descriptors. Whereas previous ML methods used expert-compiled molecular descriptors to train the algorithms, deep learning uses such features generated without any human intervention with a form of image processing called graph convolution.[48] Combining new molecular representations with multitask prediction results in models outperforming classical QSAR models.[49] To better predict molecular activities, **multitask deep learning** can also be applied to data from image analyses generated during high-content screen (HCS) assays involving the molecule itself. Such HCSs are a rich source of information, which can be used in combination with molecular descriptors to predict biological activities, while avoiding the need for customized assays.[50]

Deep learning has also been applied to *de novo* molecule generation, with the molecule being designed by the model as opposed to by the chemist. Whereas manual approaches were previously used for evolving existing molecules by adding chemical *R* groups or changing atoms, deep learning can be used to train neural networks and generate new candidates based on previously known molecules. By adapting methods commonly applied to image analysis or language translation, a first model for *de novo* molecule generation with deep learning was built up using a variational **autoencoder** encompassing both an encoder and a decoder network (Fig. 3b). The role of the encoder is to translate the chemical structure represented as a chain of characters (e.g., SMILES) into a vector called latent space. The decoder network then translates back from the latent space vectors into SMILES to obtain refined chemical structures. A random variation can be applied to the latent space or combined with model prediction to identify a decoded molecule slightly different from the input that fits the model criteria. Multiple applications of autoencoders and derivatives have been reported in combination or not with the use of recurrent neural networks (RNNs).[51–53] Additional approaches to *de novo* molecule design are being applied in computational chemistry, such as reinforcement learning (RL), in which the network is trained step by step to reach a specific output to maximize the notion of cumulative reward.[54] Another approach is to use generative adversarial networks (GAN) associating two neural networks that both compete and collaborate in a zero-sum game to perform molecular feature extraction from very large data sets. When applied to drug development, the first 'generative' network generates candidate molecules evaluated by the second 'discriminative' network.[55–57] Despite many successes obtained in drug design by using *de novo* molecule generation and multitask prediction, some of the models obtained still produce molecules that are difficult to synthesize. In this context, computational approaches have been developed to support retrosynthesis, as a substitute of expert-derived rules or knowledge-based systems built from chemical

Please cite this article in press as: P. Moingeon et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.09.006

reaction databases, by decomposing the newly generated molecule using reverse reactions to design its chemical synthesis.[58,59] Deep learning has also been recently applied to support retrosynthesis analysis using a sequence-to-sequence-based model, in which the chemical structure is described as SMILES for RNN, and the reactant and product are linked as a pair in an encoder decoder.[60] Other studies reported the use in this application of either a reaction graph[61] or a combination of three deep neural networks with a Monte Carlo tree search.[62]

Toward virtual clinical studies

AI can be used in support of the design, implementation, and monitoring of clinical trials evaluating the efficacy and safety of drug candidates, with the aim to improve success rates.[63,64] For example, the selection of patient recruited in the trials is facilitated by the understanding of disease and patient heterogeneity based on models previously discussed in the section on AI and disease modeling. In addition, NLP is being used to mine real-world evidence (RWE) data or health records to assess patient eligibility in clinical studies.[65] In this approach, automated text mining is used to identify and select patients precisely fulfilling the inclusion criteria proposed in the study design, such as level of disease severity, involvement of specific target organs, and exposure to authorized background therapies. AI is also useful to inform the design of innovative trials in a precision-medicine approach by integrating massive biological, medical imaging, and clinical data to document patient specificities. During trial monitoring, AI helps to capture in a remote fashion patient-reported measurements and outcomes generated by wearable sensors or devices. It is also applied to mine such digital biomarkers providing useful information regarding symptoms, pain, cognitive function, motricity, or sleep patterns, to support diagnostic or therapeutic decisions made by the physician.[65,66] AI and ML are also being used to analyze data from successful, but also failed studies to generate models capable of predicting simultaneously the evolution of multiple and multimodal clinical parameters.[67] Those analyses can provide hypotheses regarding candidate biomarkers predictive of progression, severity, response to treatment, or even survival in the form of genome-wide polygenic scores or multi-omics signatures.[65,68,69]

A mid-term perspective generating considerable interest is to predict the efficacy of drug candidates from virtual trials. Currently, virtual representations of the characteristics of a patient are assembled in the form of a 'synthetic' patient.[70] Those models are particularly useful as substitutes for real patients when assembling placebo control groups to test drug candidates in a life-threatening or rare disease indication. The evolution of such a virtual placebo group can be modeled from RWE clinical data obtained from real patients affected by the condition, when receiving the standard of care. Furthermore, with the aim of testing the clinical efficacy of an experimental drug, *in silico* models based on quantitative system pharmacology (QSP) are also being developed, with some encouraging results.[71,72] QSP models of a

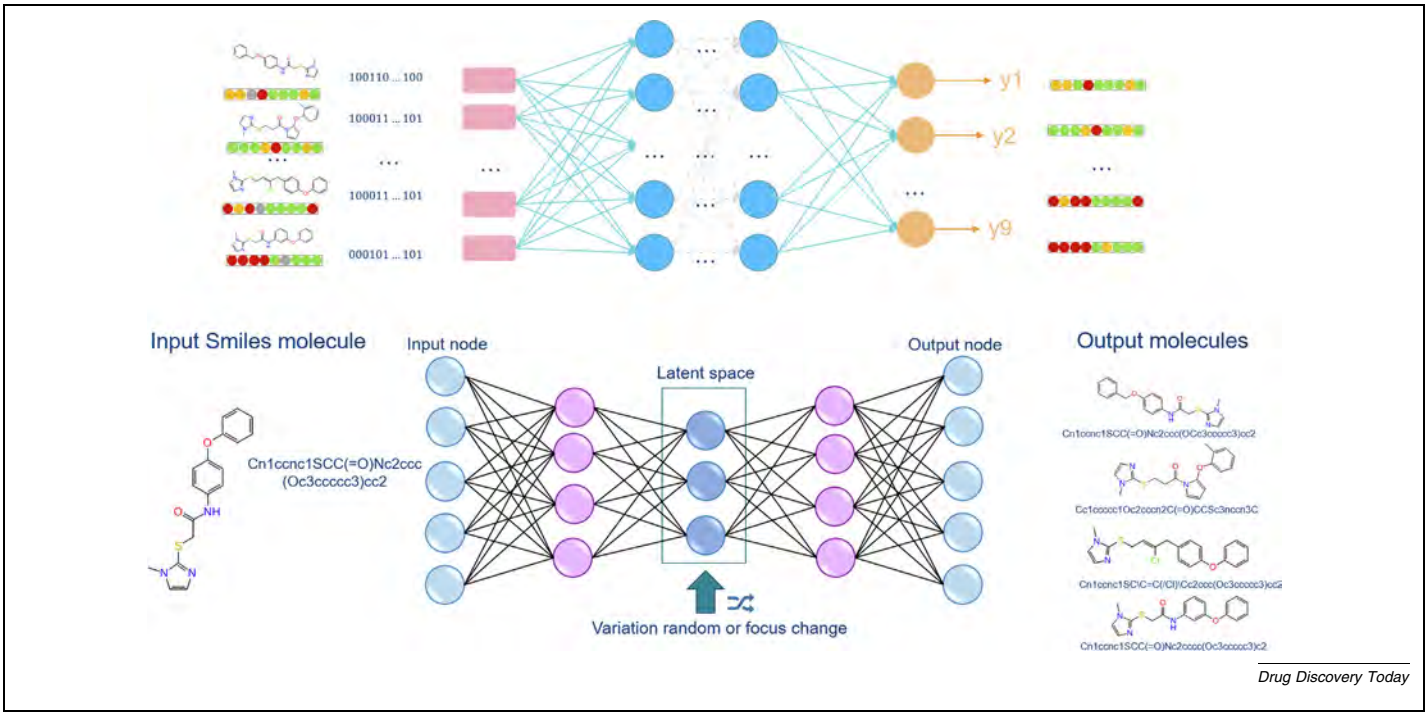

FIGURE 3

Examples of deep learning networks used in molecular modeling and drug design. **(a)** Schematic representation of a multitask prediction deep learning algorithm, on the left, with compounds used for the training and their associated data. Once the network has been trained and the best hyperparameters found, the algorithm yields as an output the full matrix prediction for all endpoints on which it has been trained. **(b)** Schematic representation of an autoencoder, with the encoder on the left, the latent space in the middle, and the decoder in the right. Once the autoencoder has been trained on millions of molecules, the latent space can be modified (through random or focus variation) to generate molecules close to the input, albeit with small changes. The autoencoder takes a SMILES as an input and produces a SMILES as the output.

Please cite this article in press as: P. Moingeon et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.09.006

disease of interest are built up from data related to biological processes in the blood or in tissues in association with clinical symptoms. The obtained biological system is then modeled as ordinary differential mathematical equations to represent dynamic interactions between components, further incorporating some main characteristics of the drug candidate (e.g., affinity for the target, pharmacokinetics, and biodistribution) to assess how the latter will perturb the system. QSP is used not only to predict how the drug could alleviate symptoms as it relates to specific organs, but also to identify potential biomarkers to categorize or monitor patients, select dosing and administration regimens as well as clinical endpoints to be used in the confirmatory real-world trial.[71,72] A remaining hurdle foreseen in implementing successfully a ML-powered precision medicine is related to the difficulty in establishing causal inferences, that is, to predict from a data-driven model a causal effect of drug exposure on clinical outcomes.[7] However, the future availability of AI-generated models of various diseases in the form of interactomes of genes or proteins with inferences of causality in the pathophysiology might considerably increase the capacity of in *silico* analyses to predict both the efficacy and safety of drug candidates.[73]

A remaining challenge to the broad application of AI to clinical studies remains the acceptance by major regulatory agencies of such virtual placebo groups, synthetic patients, and digital endpoints, as well as the validation of AI-based decision algorithms. Obviously, irrespective of advances in this field, real-world clinical studies will still be needed, likely fewer, simpler, and better designed with the help of AI.

Concluding remarks

Considering drug development as a succession of important decisions to be made to select the right target, drug, dosing regimen, and patient, it appears obvious that AI can support each of those decisions by capturing the value of massive and multimodal data into useful predictive models. Thus, AI and ML will undoubtfully produce an unprecedented revolution in drug development by making this complex and costly process ultimately cheaper and more effective, with both an anticipated shortening of the discovery phase and a reduction in failure rates during drug development. The health industry is now integrating those new technologies at a fast pace, as reflected by the exponential increase in the number of companies dedicated to AI applications to drug development (Table S2 in the supplemental information online). In 2020, a first AI-designed drug in the field of immuno-oncology entered Phase I clinical evaluation after only 12 months of research, compared with the 5–7 years commonly required in drug discovery. A new antibiotic, named halicin, has also been identified in record time using AI mining of existing molecules.[74] Numerous opportunities for drug repurposing generated by network computing have also been identified with first applications to cancers, neurological diseases, and Coronavirus 2019 (COVID-19).[4,42] Noteworthy, whereas ML has been mostly applied to the design of chemical molecules, those methods are also being considered for the design and selection of biologicals, including synthetic oligonucleotides, monoclonal antibodies, or peptides with predicted pharmacological properties.

Drug design and development encompass a range of existing human expertise, and the synergy between human and machine intelligences is vital for the successful enhancement of drug design and development. Intelligent machines can provide tremendous computing memory and power to conduct non-supervised analyses from massive multimodal data. Whereas deep learning methods are assimilated to black boxes, by contrast, humans are skilled at extracting features and providing transparency on the rationale underlying classification tasks or interpretability from the outputs of predictive models. Human expertise is needed to design and perform validation experiments in wet lab and real-world clinical studies. Importantly, human intelligence and judgement are required to consider ethical implications when implementing AI. The ultimate responsibility of diagnostic or therapeutic decisions informed by algorithms lies in healthcare professionals.

By helping to provide an unprecedented understanding of patient characteristics, AI is paving the way for a highly personalized medicine offering the perspective of future therapies and preventive measures precisely tailored to the needs of each individual patient based on their physiology and disease specificities. AI and ML also support the development of a medicine increasingly more predictive through access to multidimensional models encompassing the disease, patient, and drug candidate, and further participative by engaging patients and healthy individuals in managing their health. As such, we foresee the impact of AI and ML in the form of a rapid evolution towards an integrated computational precision medicine.

Declaration of interest

All authors are employees at SERVIER. The authors have no competing interests related to this manuscript.

Acknowledgment

The authors thank Dorothée Piva for excellent secretarial assistance.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.drudis.2021.09.006.

References

1 J.A. DiMasi, H.G. Grabowski, R.W. Hansen, Innovation in the pharmaceutical industry: new estimates of R&D costs, J Health Econ 47 (2016) 20–33.

2 M.J. Waring, J. Arrowsmith, A.R. Leach, P.D. Leeson, S. Mandrell, R.M. Owen, G. Pairaudeau, W.D. Pennie, S.D. Pickett, J. Wang, O. Wallace, A. Weir, An analysis of the attrition of drug candidates from four major pharmaceutical companies, Nat Rev Drug Discov 14 (7) (2015) 475–486.

3 K.-K. Mak, M.R. Pichika, Artificial intelligence in drug development: present status and future prospects, Drug Discovery Today 24 (3) (2019) 773–780.

4 J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer, S. Zhao, Applications of machine learning in drug discovery and development, Nature Reviews Drug Discovery 18 (6) (2019) 463–477.

Please cite this article in press as: P. Moingeon et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.09.006

5 D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, R.K. Tekade, Artificial intelligence in drug discovery and development, Drug Discov Today 26 (1) (2021) 80–93.
6 J. Xu, P. Yang, S. Xue, B. Sharma, M. Sanchez-Martin, F. Wang, K.A. Beaty, E. Dehan, B. Parikh, Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives, Hum Genet 138 (2) (2019) 109–124.
7 J. Wilkinson, K.F. Arnold, E.J. Murray, M. van Smeden, K. Carr, R. Sippy, M. de Kamps, A. Beam, S. Konigorski, C. Lippert, M.S. Gilthorpe, P.W.G. Tennant, Time to reality check the promises of machine learning-powered precision medicine, The Lancet Digital Health 2 (12) (2020) e677–e680.
8 CDISC | Clear Data. Clear Impact. https://www.cdisc.org/ [Accessed September 14, 2021].
9 FAIR Principles. GO FAIR. https://www.go-fair.org/fair-principles/ [Accessed September 14, 2021].
10 M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3 (1) (2016), https://doi.org/10.1038/sdata.2016.18.
11 General Data Protection Regulation (GDPR) – Official Legal Text. General Data Protection Regulation (GDPR). https: //gdpr-info.eu/ [Accessed September 14, 2021].
12 MELLODDY. https://www.melloddy.eu [Accessed September 14, 2021].
13 J. Tang, Z.-u.-R. Tanoli, B. Ravikumar, Z. Alam, A. Rebane, M. Vähä-Koskela, G. Peddinti, A.J. van Adrichem, J. Wakkinen, A. Jaiswal, E. Karjalainen, P. Gautam, L. He, E. Parri, S. Khan, A. Gupta, M. Ali, L. Yetukuri, A.-L. Gustavsson, B. Seashore-Ludlow, A. Hersey, A.R. Leach, J.P. Overington, G. Repasky, K. Wennerberg, T. Aittokallio, Drug target commons: a community effort to build a consensus knowledge base for drug-target interactions, Cell Chem Biol 25 (2) (2018) 224–229.e2.
14 Kaggle: Your Machine Learning and Data Science Community. https://www.kaggle.com/ [Accessed July 12, 2021].
15 DREAM Challenges. https://dreamchallenges.org/ [Accessed September 14, 2021].
16 PrecisionFDA Truth Challenge – precisionFDA. https://precision.fda.gov/challenges/truth [Accessed September 14, 2021].
17 E.P. García del Valle, G. Lagunes García, L. Prieto Santamaría, M. Zanin, E. Menasalvas Ruiz, A. Rodríguez-González, Disease networks and their contribution to disease understanding: a review of their evolution, techniques and data sources, J Biomed Inform 94 (2019) 103206, https://doi.org/10.1016/j.jbi.2019.103206.
18 S.A. Dugger, A. Platt, D.B. Goldstein, Drug development in the era of precision medicine, Nat Rev Drug Discov 17 (3) (2018) 183–196.
19 L.-H. Lee, J. Loscalzo, Network medicine in pathobiology, Am J Pathol 189 (7) (2019) 1311–1326.
20 Y. Hasin, M. Seldin, A. Lusis, Multi-omics approaches to disease, Genome Biol 18 (1) (2017) 83.
21 G. Tini, L. Marchetti, C. Priami, M.-P. Scott-Boyer, Multi-omics integration-a comparison of unsupervised clustering methodologies, Brief Bioinform 20 (4) (2019) 1269–1279.
22 E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I.W.H. Kwok, L.G. Ng, F. Ginhoux, E.W. Newell, Dimensionality reduction for visualizing single-cell data using UMAP, Nature Biotechnology 37 (1) (2019) 38–44.
23 I.C. Macaulay, C.P. Ponting, T. Voet, Single-cell multiomics: multiple measurements from single cells, Trends Genet 33 (2) (2017) 155–168.
24 J.C.B. Gamboa, Deep learning for time-series analysis, arXiv 2017 (1887) 17010.
25 J. Guinney, R. Dienstmann, X. Wang, A. de Reyniès, A. Schlicker, C. Soneson, L. Marisa, P. Roepman, G. Nyamundanda, P. Angelino, B.M. Bot, J.S. Morris, I.M. Simon, S. Gerster, E. Fessler, F. De Sousa E Melo, E. Missiaglia, H. Ramay, D. Barras, K. Homicsko, D. Maru, G.C. Manyam, B. Broom, V. Boige, B. Perez-Villamil, T. Laderas, R. Salazar, J.W. Gray, D. Hanahan, J. Tabernero, R. Bernards, S.H. Friend, P. Laurent-Puig, J.P. Medema, A. Sadanandam, L. Wessels, M. Delorenzi, S. Kopetz, L. Vermeulen, S. Tejpar, The consensus molecular subtypes of colorectal cancer, Nature Medicine 21 (11) (2015) 1350–1356.
26 C. Ai, L. Kong, CGPS: a machine learning-based approach integrating multiple gene set analysis tools for better prioritization of biologically relevant pathways, Journal of Genetics and Genomics 45 (9) (2018) 489–504.
27 U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F.H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksöz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, E.E. Wanker, A human protein-protein interaction network: a resource for annotating the proteome, Cell 122 (6) (2005) 957–968.
28 N. Friedman, Inferring cellular networks using probabilistic graphical models, Science 303 (5659) (2004) 799–805.
29 C.J. Needham, J.R. Bradford, A.J. Bulpitt, D.R. Westhead, Inference in Bayesian networks, Nature Biotechnology 24 (1) (2006) 51–53.
30 A.-L. Barabási, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease, Nature Reviews Genetics 12 (1) (2011) 56–68.
31 L. Cowen, T. Ideker, B.J. Raphael, R. Sharan, Network propagation: a universal amplifier of genetic associations, Nature Reviews Genetics 18 (9) (2017) 551–562.
32 L. Cantini, E. Medico, S. Fortunato, M. Caselle, Detection of gene communities in multi-networks reveals cancer drivers, Scientific Reports 5 (1) (2015) 17386.
33 Raue A, Schilling M, Bachmann J, Matteson A, Schelker M, Kaschek D, et al. Lessons learned from quantitative dynamical modeling in systems biology. PLoS ONE 2013; 8(9): e74335.
34 Ochoa D, Hercules A, Carmona M, Suveges D, Gonzalez-Uriarte A, Malangone C, et al. Open Targets Platform: supporting systematic drug–target identification and prioritisation. Nucleic Acids Research 2021; 49(D1): D1302-D1310.
35 Owens J. Determining druggability. Nature Reviews Drug Discovery 2007; 6(3): 187-187.
36 W. Muster, A. Breidenbach, H. Fischer, S. Kirchner, L. Müller, A. Pähler, Computational toxicology in drug development, Drug Discovery Today 13 (7-8) (2008) 303–310.
37 L.J. Jensen, J. Saric, P. Bork, Literature mining for the biologist: from information retrieval to biological discovery, Nature Reviews Genetics 7 (2) (2006) 119–129.
38 V. Vergetis, D. Skaltsas, V.G. Gorgoulis, A. Tsirigos, Assessing drug development risk using big data and machine learning, Cancer Res 81 (4) (2021) 816–819.
39 Skalic M, Varela-Rial A, Jiménez J, Martínez-Rosell G, De Fabritiis G. LigVoxel: inpainting binding pockets using 3D-convolutional neural networks. Bioinformatics 2019; 35(2): 243-250.
40 A.W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A.W.R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D.T. Jones, D. Silver, K. Kavukcuoglu, D. Hassabis, Improved protein structure prediction using potentials from deep learning, Nature 577 (7792) (2020) 706–710.
41 E. Guney, J. Menche, M. Vidal, A.-L. Barábasi, Network-based in silico drug efficacy screening, Nature Communications 7 (1) (2016) 10331.
42 F. Cheng, R.J. Desai, D.E. Handy, R. Wang, S. Schneeweiss, A.-L. Baraba´si, J. Loscalzo, Network-based approach to prediction and population-based validation of in silico drug repurposing, Nat Commun 9 (1) (2018), https://doi.org/10.1038/s41467-018-05116-5.
43 X. Zeng, S. Zhu, W. Lu, Z. Liu, J. Huang, Y. Zhou, J. Fang, Y. Huang, H. Guo, L. Li, B.D. Trapp, R. Nussinov, C. Eng, J. Loscalzo, F. Cheng, Target identification among known drugs by deep learning from heterogeneous networks, Chem Sci 11 (7) (2020) 1775–1797.
44 Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014; 15(1): 1929-1958.]
45 Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: Fürnkranz J, Joachims T, eds. Proceedings of the 27th International Conference on International Conference on Machine Learning. ICML'10. Madison: Omnipress, 2010: 807-814
46 A.S. Rifaioglu, H. Atas, M.J. Martin, R. Cetin-Atalay, V. Atalay, T. Doğan, Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases, Brief Bioinform 20 (5) (2019) 1878–1912.
47 B. Ramsundar, B. Liu, Z. Wu, A. Verras, M. Tudor, R.P. Sheridan, V. Pande, Is multitask deep learning practical for pharma?, J Chem Inf Model 57 (8) (2017) 2068–2076
48 Wieder O, Kohlbacher S, Kuenemann M, Garona A, Ducrota P, Seidel T et al. A compact review of molecular property prediction with graph neural networks. Drug Discovery Today: Technologies. Published online December 17, 2020. http://dx.doi.org/10.1016/j.ddtec.2020.11.009
49 Walters WP, Barzilay R. Applications of deep learning in molecule generation and molecular property prediction. Acc Chem Res 2021; 54(2): 263-270.

Please cite this article in press as: P. Moingeon et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.09.006

50 J. Simm, G. Klambauer, A. Arany, M. Steijaert, J.K. Wegner, E. Gustin, V. Chupakhin, Y.T. Chong, J. Vialard, P. Buijnsters, I. Velter, A. Vapirev, S. Singh, A. E. Carpenter, R. Wuyts, S. Hochreiter, Y. Moreau, H. Ceulemans, Repurposing high-throughput image assays enables biological activity prediction for drug discovery, Cell Chem Biol 25 (5) (2018) 611–618.e3.

51 A. Kadurin, S. Nikolenko, K. Khrabrov, A. Aliper, A. Zhavoronkov, druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico, Mol Pharm 14 (9) (2017) 3098–3104.

52 T. Blaschke, M. Olivecrona, O. Engkvist, J. Bajorath, H. Chen, Application of generative autoencoder in de novo molecular design, Mol Inform 37 (1-2) (2018) 1700123, https://doi.org/10.1002/minf.201700123.

53 M.H.S. Segler, T. Kogej, C. Tyrchan, M.P. Waller, Generating focused molecule libraries for drug discovery with recurrent neural networks, ACS Cent Sci 4 (1) (2018) 120–131.

54 M. Olivecrona, T. Blaschke, O. Engkvist, H. Chen, Molecular de-novo design through deep reinforcement learning, J Cheminform 9 (1) (2017) 48.

55 Guimaraes GL, Sanchez-Lengeling B, Outeiral C, Farias PLC, Aspuru-Guzik A. Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models. arXiv 2018: 170510843.

56 De Cao N, Kipf T. MolGAN: an implicit generative model for small molecular graphs. arXiv 2018: 180511973.

57 E. Putin, A. Asadulaev, Y. Ivanenkov, V. Aladinskiy, B. Sanchez-Lengeling, A. Aspuru-Guzik, A. Zhavoronkov, Reinforced adversarial neural computer for de novo molecular design, J Chem Inf Model 58 (6) (2018) 1194–1204.

58 O. Engkvist, P.-O. Norrby, N. Selmi, Y.-H. Lam, Z. Peng, E.C. Sherer, W. Amberg, T. Erhard, L.A. Smyth, Computational prediction of chemical reactions: current status and outlook, Drug Discov Today 23 (6) (2018) 1203–1218.

59 J. Law, Z. Zsoldos, A. Simon, D. Reid, Y. Liu, S.Y. Khew, A.P. Johnson, S. Major, R. A. Wade, H.Y. Ando, Route Designer: a retrosynthetic analysis tool utilizing automated retrosynthetic rule generation, J Chem Inf Model 49 (3) (2009) 593–602.

60 B. Liu, B. Ramsundar, P. Kawthekar, J. Shi, J. Gomes, Q. Luu Nguyen, S. Ho, J. Sloane, P. Wender, V. Pande, Retrosynthetic reaction prediction using neural sequence-to-sequence models, ACS Cent Sci 3 (10) (2017) 1103–1113.

61 Savage J, Kishimoto A, Buesser B, Diaz-Aviles E, Alzate C. Chemical reactant recommendation using a network of organic chemistry. In: Cremonesi P, Ricci F, eds; RecSys '17: Proceedings of the Eleventh ACM Conference on Recommender Systems. New York: Association for Computing Machinery, 2017: 210-214.

62 M.H.S. Segler, M. Preuss, M.P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI, Nature 555 (7698) (2018) 604–610.

63 M. Hay, D.W. Thomas, J.L. Craighead, C. Economides, J. Rosenthal, Clinical development success rates for investigational drugs, Nat Biotechnol 32 (1) (2014) 40–51.

64 S. Harrer, P. Shah, B. Antony, J. Hu, Artificial intelligence for clinical trial design, Trends in Pharmacological Sciences 40 (8) (2019) 577–591.

65 P. Shah, F. Kendall, S. Khozin, R. Goosen, J. Hu, J. Laramie, M. Ringel, N. Schork, Artificial intelligence and machine learning in clinical development: a translational perspective, NPJ Digital Medicine 2 (1) (2019), https://doi.org/10.1038/s41746-019-0148-3.

66 P. Boehme, A. Hansen, R. Roubenoff, J. Scheeren, M. Herrmann, T. Mondritzki, J. Ehlers, H. Truebel, How soon will digital endpoints become a cornerstone for future drug development?, Drug Discov Today 24 (1) (2019) 16–19

67 D.B. Fogel, Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review, Contemp Clin Trials Commun 11 (2018) 156–164.

68 A.V. Khera, M. Chaffin, K.G. Aragam, M.E. Haas, C. Roselli, S.H. Choi, P. Natarajan, E.S. Lander, S.A. Lubitz, P.T. Ellinor, S. Kathiresan, Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations, Nat Genet 50 (9) (2018) 1219–1224.

69 A.J. Steele, S.C. Denaxas, A.D. Shah, H. Hemingway, N.M. Luscombe, T.R. Singh, Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease, PLoS ONE 13 (8) (2018) e0202344, https://doi.org/10.1371/journal.pone.020234410.1371/journal.pone.0202344.g00110.1371/journal.pone.0202344.g00210.1371/journal.pone.0202344.g00310.1371/journal.pone.0202344.g00410.1371/journal.pone.0202344.g00510.1371/journal.pone.0202344.t00110.1371/journal.pone.0202344.t002.

70 A. Tucker, Z. Wang, Y. Rotalinti, P. Myles, Generating high-fidelity synthetic patient data for assessing machine learning healthcare software, NPJ Digital Medicine 3 (1) (2020) 1–13.

71 Sorger P, Allerheiligen SRB. Quantitative and Systems Pharmacology in the Post-genomic Era: New Approaches to Discovering Drugs and Understanding Therapeutic Mechanisms. 2011. https://www.nigms.nih.gov/training/documents/systemspharmawpsorger2011.pdf [Accessed September 14, 2021]

72 C.M. Friedrich, A model qualification method for mechanistic physiological QSP models to support model-informed drug development, CPT Pharmacometrics Syst Pharmacol 5 (2) (2016) 43–53.

73 M. Danhof, Systems pharmacology - towards the modeling of network interactions, Eur J Pharm Sci 94 (2016) 4–14.

74 J.M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N.M. Donghia, C.R. MacNair, S. French, L.A. Carfrae, Z. Bloom-Ackermann, V.M. Tran, A. Chiappino-Pepe, A.H. Badran, I.W. Andrews, E.J. Chory, G.M. Church, E.D. Brown, T.S. Jaakkola, R. Barzilay, J.J. Collins, A deep learning approach to antibiotic discovery, Cell 180 (4) (2020) 688–702.e13.

Glossary

Artificial Intelligence (AI): any type of machine presenting a simple intelligence. In computer science, AI is defined as a machine able to perform tasks requiring human intelligence, such as visual perception, speech recognition, decision-making, and language translation.

Autoencoder: neural network technique that performs dimensionality reduction. It comprises an encoder part compressing and encoding data efficiently, and a decoder part, which learns how to reconstruct the data back as close as possible to the original.

Consensus: convergence between predictions obtained from different models, each of them generated from either different data sources (e.g., multi-omics) or computational approaches (e.g., hierarchical, Gaussian, and k-means clusterings).

Data sparsity: computational challenge linked to the completeness of the observations in a data set.

Deep learning: advanced algorithm mimicking the human brain by using artificial neurons in a complex network.

Gradient problem: a gradient measures how much the output of a model changes when inputs are modified. In gradient-based approaches, such as neural networks, gradients are used during training to update some parameter weights. When the magnitudes of the gradients accumulate, an unstable model is likely to occur, which can lead to poor prediction results. Methods to manage exploding gradients include gradient clipping and weight regularization.

Layer: structure in the architecture of the model, which enables information to be taken from a previous layer and be passed on to the next one. Different types of layer exist, such as fully connected or convolutional layers.

Machine learning (ML): use and development of algorithms based on sample data (e.g., experimental data), which can learn and adapt without human instructions to analyze and draw inferences from patterns in data.

Multitask deep learning: process to solve multiple learning tasks simultaneously, while exploiting commonalities and differences across each of them. Multitasking has been used successfully across all applications of ML.

Natural language processing (NLP): treatment of human language by intelligent machines to understand and extract relevant information from the content of data sources, such as publications or patents.

Networks/knowledge graphs: set of entities or nodes (e.g., genes, proteins, drugs, or diseases), connected to each other by relationships or links (e.g., gene–disease association, protein–protein interaction, or drug–target interactions).

Network propagation: algorithms (e.g., random walk or information diffusion) used to propagate data into the topology of a given network to amplify colocalized high signals and support functional interpretation.

Overfitting: predictive model corresponding too closely to the data set on which it has been trained, which therefore might fail to be validated in other data sets. Cross-validation or bootstrapping resampling approaches are generally proposed to reduce the overfitting effect.

Supervised learning: aims to create a prediction function based on labeled data (as opposed to unsupervised learning); encompasses both classification learning based on qualitative data and regression learning trained by using quantitative data.

Unsupervised learning: a mode of ML in which data are not labeled; aims to discove r the underlying structures to label or group those unlabeled data.

Please cite this article in press as: P. Moingeon et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.09.006

Network-based repurposing identifies anti-alarmins as drug candidates to control severe lung inflammation in COVID-19

Emiko Desvaux , Antoine Hamon , Sandra Hubert , Cheïma Boudjeniba, Bastien Chassagnol, Jack Swindle, Audrey Aussy, Laurence Laigle, Jessica Laplume, Perrine Soret, Pierre Jean-François, Isabelle Dupin-Roger, Mickaël Guedj , Philippe Moingeon

Published: July 22, 2021 • https://doi.org/10.1371/journal.pone.0254374

Abstract

While establishing worldwide collective immunity with anti SARS-CoV-2 vaccines, COVID-19 remains a major health issue with dramatic ensuing economic consequences. In the transition, repurposing existing drugs remains the fastest cost-effective approach to alleviate the burden on health services, most particularly by reducing the incidence of the acute respiratory distress syndrome associated with severe COVID-19. We undertook a computational repurposing approach to identify candidate therapeutic drugs to control progression towards severe airways inflammation during COVID-19. Molecular profiling data were obtained from public sources regarding SARS-CoV-2 infected epithelial or endothelial cells, immune dysregulations associated with severe COVID-19 and lung inflammation induced by other respiratory viruses. From these data, we generated a protein-protein interactome modeling the evolution of lung inflammation during COVID-19 from inception to an established cytokine release syndrome. This predictive model assembling severe COVID-19-related proteins supports a role for known contributors to the cytokine storm such as IL1β, IL6, TNFα, JAK2, but also less prominent actors such as IL17, IL23 and C5a. Importantly our analysis points out to alarmins such as TSLP, IL33, members of the S100 family and their receptors (ST2, RAGE) as targets of major therapeutic interest. By evaluating the network-based distances between severe COVID-19-related proteins and known drug targets, network computing identified drugs which could be repurposed to prevent or slow down progression towards severe airways inflammation. This analysis confirmed the interest of dexamethasone, JAK2 inhibitors, estrogens and further identified various drugs either available or in development interacting with the aforementioned targets. We most particularly recommend considering various inhibitors of alarmins or their receptors, currently receiving little attention in this indication, as candidate treatments for severe COVID-19.

Citation: Desvaux E, Hamon A, Hubert S, Boudjeniba C, Chassagnol B, Swindle J, et al. (2021) Network-based repurposing identifies anti-alarmins as drug candidates to control severe lung inflammation in COVID-19. PLoS ONE 16(7): e0254374. https://doi.org/10.1371/journal.pone.0254374

Editor: Svetlana P. Chapoval, University of Maryland School of Medicine, UNITED STATES

Received: March 5, 2021; **Accepted:** June 24, 2021; **Published:** July 22, 2021

Data Availability: The RNA Seq data on genes differentially expressed in SARS-CoV-2 infected NHBE or Calu-3 human lung epithelial cells are publicly available from repository Gene Expression Omnibus (GEO, accession number GSE147507). Drug-target links can be retrieved from the Therapeutic Target Database (version 7.1.01) and from Drugbank. All other sources of data used in the present study related to various aspects of COVID-19 pathophysiology were obtained from the scientific literature. A comprehensive list of those publications and detailed references are provided in Supporting information file S1 Table.

Funding: We confirm that Servier and Lincoln only provided financial support in the form of authors' salaries. These companies did not play a role in the study design, data collection and analysis, decision to publish, nor in the preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: We confirm that our commercial affiliations do not alter our adherence to PLOS ONE policies on sharing data and materials.

Introduction

Since the emergence of the new strain of Coronavirus SARS-CoV-2 in December 2019, the ongoing crisis associated with the COVID-19 disease has affected more than 170 million individuals worldwide, causing over 3.5 million deaths (World Health Organization Dashboard, June 1st, 2021), mainly as the consequence of an Acute Respiratory Distress Syndrome (ARDS). The pandemic is still progressing actively despite lockdown measures throughout the world, with the recent emergence of highly transmissible viral strains [1]. To date, the only proven medications for reducing either viral loads, hospitalization rates, invasive mechanical ventilation or patient mortality include corticosteroids such as dexamethasone, the antiviral remdesivir, the anti-IL6R tocilizumab as well as neutralizing monoclonal antibodies directed to the spike protein of the virus [2–5]. Many additional drugs have been tested, including the lopinavir antiviral, the anti-malarial hydroxychloroquine or IFNβ with as of today disappointing efficacy results [6].

Recently, several vaccines have been approved by regulatory authorities based on remarkable efficacy results, with evidence that they can protect against infection by eliciting high titers of neutralizing antibodies against the Spike protein of the SARS-CoV-2 virus [7]. Whereas such vaccines will very positively transform the course and gravity of the COVID-19 pandemic, a recent concern is whether they will be fully effective against emerging new variants of the virus bearing point mutations in the Spike protein [1]. Furthermore, the challenge of manufacturing and administering billions of vaccine doses in order to establish a protective herd immunity at a worldwide population level will not be met in a short time frame.

During the time needed to deploy preventive vaccines at such a scale, the repurposing of existing drugs is a valid solution to better address severe forms of COVID-19 and alleviate the burden on health services in a time and cost-effective manner. Previous repurposing strategies have been undertaken in the context of a limited understanding of COVID-19 pathogenesis, prompting to use related viruses such as SARS-CoV and MERS-CoV as proxies to model SARS-CoV-2 infection [8–13]. Several network computing studies have been successful to predict drug disease associations for repurposing in COVID-19. Many of those initial approaches were aiming to identify existing compounds to prevent viral infection by either targeting mechanisms involving the viral receptor ACE2 (angiotensin converting enzyme 2), the TMPRSS2 transmembrane protease serine 2, or clathrin-mediated endocytosis [14–16]. In the present repurposing study, we rather focused on drugs predicted to interfere with pro-inflammatory mediators identified by modelling immune dysregulations caused in the airways by SARS-CoV-2 infection.

Since a vast majority of patients infected with SARS-CoV-2 develop no or only mild symptoms, we reasoned that ideal candidate drugs to repurpose should rather inhibit severe airways inflammation in the course of the disease. Lung inflammation is the main cause requiring hospitalization in up to 20% of COVID-19 cases, with life threatening ARDS affecting 75% of COVID-19 patients transferred to intensive care units [17]. In this subset of patients with severe lung inflammation, persisting proinflammatory immune responses result in a cytokine release syndrome (CRS) linked to the activation of myeloid cells secreting cytokines such as IL1β, IL6 and TNFα [18–20].

Capitalizing on the most recent scientific insights on the pathophysiology of COVID-19, we undertook computational network analyses to integrate a wide variety of data sources encompassing extensive molecular profiling of SARS-CoV-2 infected epithelial or endothelial cells, genetic susceptibilities and immune dysregulations linked to severe COVID-19 as well as molecular mechanisms elicited during lung infection by other respiratory viruses. From this approach, a short list of COVID-19 disease-related proteins considered as potential therapeutic targets was established and used to computationally assess a topological proximity with drug targets within the comprehensive human protein-protein interactome [21, 22]. Herein, we report on the identification of candidate therapeutic targets, as well as drugs predicted to interact with some of those targets which could be repurposed to prevent or slow down severe lung inflammation during COVID-19.

Materials and methods

Sources of data on COVID-19 pathophysiology

To identify proteins related to lung inflammation in COVID-19, we selected relevant categories of data from the scientific literature (detailed in S1 Table), such as genes differentially expressed following SARS-CoV-2 infection of (i) primary normal human bronchial epithelial cells (*NHBE*) or of the ACE2-expressing lung-epithelial *Calu-3* cell line, (ii) endothelial cells or cells recovered from bronchoalveolar lavages or lung biopsies of patients with severe COVID-19 [23–25]. We also mined public data regarding immunological signatures obtained in the blood or in tissues of patients, distinguishing those with mild COVID-19 from others rather affected by severe forms of the disease [26–34]. We included as well information from previous studies on lung inflammation caused by other respiratory viruses (including asthma exacerbation), in light of an involvement of monocytes, macrophages, myeloid dendritic cells, innate lymphoid cells in those conditions similarly to COVID-19 [18, 35–38].

Identification of disease-related proteins

COVID-19 disease-related proteins predicted to be involved in early lung inflammation and in the transition to the cytokine storm were identified following data mining from scientific publications listed in S1 Table. To establish molecular pathways dysregulated during lung inflammation due to COVID-19, we first used RNAseq data from *NHBE* (normal human bronchial epithelial) and *Calu-3* (human lung epithelial cancer) cells infected or not with SARS-CoV-2. These data were pre-treated by removing outlier samples whose total sum of counts was below 5 000 000. In order to filter out genes undistinguishable from background noise, we modelled gene expression after applying a log2(x + 1) transformation by a two component Gaussian mixture model, with a first peak corresponding to unexpressed genes, and the second peak to truly expressed genes. Numbers of genes pre and post-filtering were 17557 and 21797, respectively. We retrieved the parameters of the mixture distribution using function normalmixEM from mixtools package and determined that the 0.95 quantile for the noise distribution was 1.6. We subsequently removed all genes whose expression was below that threshold in more than 95% of samples. We performed a differential analysis (COVID versus mock) in each cell line using the *limma R* package and *eBayes* function (with mock group corresponding to healthy & no treatment patients). Disease signatures were then extracted by considering differentially expressed genes (DEG) as those with adjusted *p*-value below 0.05 with an absolute fold change superior to 1.3 (commonly used as a threshold for biological significance). Canonical pathway enrichment analyses were subsequently performed by using the Ingenuity Pathway Analysis (IPA) software.

Network-based drug repurposing

Network-based drug repurposing relies on the hypothesis that the closer a target is to a group of disease related genes in the PPI network, the higher the chance of having a significant impact on the disease. Many approaches focus on the shortest path to determine proximity, with some variations in order to avoid hub protein bias [15, 39]. The latter bias occurs from certain proteins that have an extremely high degree in the network and thereby cause a highly dense graph. Other approaches take advantage of the diffusion process to define proximity [40] while considering all the topological features of the graph. Diffusion based metrics have a comparable advantage over shortest path distances when in highly dense graphs such as PPI graphs [41]. Other metrics distinct from shortest path and diffusion can be used such as such as largest connected component -based methods [42].

Our computational repurposing approach (Fig 1A) takes advantage of the proximity between disease-related proteins and drug targets through an established network of protein-protein interactions (PPIs, referred to as an *interactome*). Drug-target links were gathered from the Therapeutic Target Database (TTD, version 7.1.01) and Drugbank [43, 44]. The PPIs network was derived from previous work by Cheng et al [45]. It was built from 15 different databases such as BioGRID and HPRD by compiling binary PPIs tested by high-throughput yeast-two-hybrid (Y2H) systems, kinase-substrate interactions from literature-derived low-throughput and high-throughput experiments, high-quality PPIs from three-dimensional (3D) protein structures, and signaling networks from literature-derived low-throughput experiments.

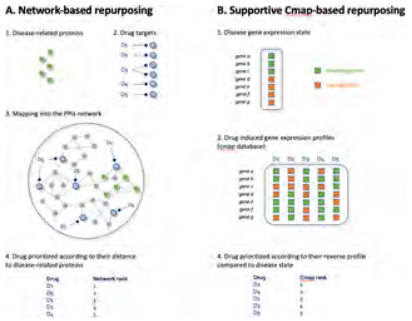

Fig 1. General principles of network and Cmap-based repurposing approaches.
A) Network-based repurposing. Disease-related proteins and drug targets are mapped into a network of protein-protein-interactions (PPI). Drugs are prioritized according to their distance to disease-related proteins. B) Supportive Cmap-based repurposing. In those supportive analyses, disease-related as well as drug induced gene expression states are compared in order to identify drugs eliciting reverse profiles compared to those found in the disease.
https://doi.org/10.1371/journal.pone.0254374.g001

Relevance of drugs to the disease was assessed based on proximity of their targets to disease-related proteins according to two complementary metrics, namely a simple *topological distance* and a more advanced *diffusion-based distance*.

The *topological distance* (d_{topo}) corresponds to the shortest path length in the PPIs network between the disease-related proteins and the drug targets, computed according to the following formula:

$$d_{\text{topo}}(P,T) = \frac{1}{\|T\|}\sum_{t\in T}\min_{p\in P} SP(p,t)$$

With *P* the set of nodes corresponding to the disease-related proteins, *T* the set of nodes corresponding to the drug targets, and *SP*(*p*,*t*) the shortest path length between a node *p* of *P* and another node *t* of *T*. When calculating a topological distance, we generate a distribution from bootstrapping similar nodes defined by same degree in the graph. From the given distribution, we calculate a z-score (and p-value).

The *diffusion-based distance* (d_{diff}) is computed based on the similarity of the impact on the network of perturbations starting from disease-related proteins on one side and drug targets on the other. The impact of a perturbation starting from a given node n_i on the network is assessed by use of a *diffusion* algorithm. Let (n_i,n_j) being a pair of nodes, then $\mathbb{P}(n_i, n_j)$ represents the random walk-based probability that a perturbation starting from n_i reaches n_j. It allows us to define a numerical vector $V(n_i)$ representing the impact perturbation of n_i on the whole interactome:

$$V(n_i) = [\mathbb{P}(n_i,\ n_1),\ \mathbb{P}(n_i,\ n_2),\ \ldots,\ \mathbb{P}(n_i,n_n)]$$

The similarity between two perturbations starting from n_i and n_j is then assessed by computing the Manhattan distance between $V(n_i)$ and $V(n_j)$. In order to extend this principle to the distance between sets of nodes, we derived the following formula:

$$d_{\text{diff}}(P,T) = \frac{1}{\|T\|}\sum_{t\in T}\min_{p\in P} MD(p,t)$$

With *P* the set of nodes corresponding to the disease-related proteins, *T* the set of nodes corresponding to the drug targets, *p* one given node of *P*, *t* one given node of *T*, and *MD* (*p*,*t*) the Manhattan distance between *V*(*p*) and *V*(*t*). This diffusion-based distance was implemented via the DSD algorithm [46]. For each diffusion-based distance, we also calculate associated z-scores (and p-values). Note that DSD is by construction normally distributed. In order to prioritize drugs from this network-based repurposing approach, we defined a network rank resulting from the mean rank aggregation of d_{topo} and d_{diff}. Given that we have p-values for both of our distance measures, we perform a Fisher's combined probability test to obtain a unique combined p-value per drug. Using the DSD algorithm, we generated a computed distance matrix of 15 894 X 15 894 encompassing all proteins in our interactome.

Cmap-based drug repurposing

We complemented the network-based approach by using Cmap as a supportive method (Fig 1B). Cmap identifies drugs inducing a reverse gene expression profile compared to the disease state using a method of similarity [47]. The Cmap database comprises human cancer cell lines either treated or not with chemical drugs, referred to as perturbagens. We used the *R* package *ccdata* which encompasses expression profiles for 1309 perturbagens over 13832 genes. Disease state was obtained from gene expression profiles induced in *NHBE* and *Calu-3* cells following infection by SARS-CoV-2. We compare expression profiles induced by disease state with those induced by perturbagens, using mainly the Pearson correlation between transcriptome values of the query signature and the perturbagen signature. A negative correlation score provides a potential therapeutic indication of the perturbagen. Cmap scores (the smaller the better) were first computed on both *NHBE* and C*alu-3* data and then averaged.

Results and discussion

Identification of COVID-19 disease-related proteins

Based on recent scientific advances, the pathophysiology of COVID-19 can be summarized as three sequential steps (Fig 2). We reasoned that treatments suitable to control severe COVID-19 should interfere with molecular pathways involved in the evolution from mild to severe lung inflammation (Fig 2, central panel), while preserving anti-viral protective immune mechanisms. We thus compiled a comprehensive list of genes differentially upregulated in *NHBE* and *Calu-3* human epithelial cells following SARS-CoV-2 infection, providing important quantitative information [23]. We cross-validated this list in comparison with molecular signatures

reported at the level of endothelial cells, bronchoalveolar lavage cells or lung biopsies in other studies to be associated with severe COVID-19 or exposure to other respiratory viruses (S1 Table). The latter was further completed with deep immunophenotyping, RNA seq and cytokine profiling data related to dysregulated innate or adaptive immune responses in the blood or the lungs of patients with severe COVID-19. A compilation of the most relevant COVID-19 disease related-proteins thus obtained, together with data sources supporting their relevance to lung inflammation in COVID-19 are presented in S1 Table.

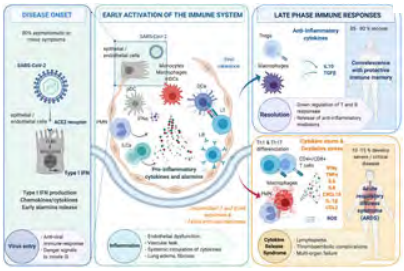

Fig 2. Three step progression towards severe COVID-19.
The pathophysiology of COVID-19 in the airways encompasses schematically three successive steps, including **(i)** Disease onset following viral infection of alveolar epithelial or endothelial cells expressing the ACE2 receptor (left panel) leading to the activation of the innate immune system, with IFNα production by plasmacytoid dendritic cells (pDC). **(ii)** An early inflammatory phase within lung tissues where a cross-talk between infected epithelial/endothelial cells and innate immune cells such as monocytes, macrophages, myeloid dendritic cells (mDC) and innate lymphoid cells (ILCs) leads to a release of pro-inflammatory alarmins, cytokines and chemokines (center panel). This results in the activation of adaptive immunity, involving both CD4+ T cell help, CD8+ T cells cytotoxic for virally-infected cells as well as production of neutralizing antibodies against surface viral antigens. **(iii)** A late inflammatory phase with two potential outcomes: 85 to 90% of cases evolve towards resolution of inflammation with downregulation of T and B cell responses concomitant with the release of anti-inflammatory mediators (right upper panel); whereas 10 to 15% patients rather exhibit major tissue damage and severe acute respiratory distress syndrome (ARDS) caused by a deleterious uncontrolled inflammation linked with persisting T cell activation, excessive myeloid cell activation associated with a cytokine storm as well as oxidative stress (right lower panel).
https://doi.org/10.1371/journal.pone.0254374.g002

Ingenuity pathway analyses were then performed on this list, allowing to confirm that genes/proteins upregulated following SARS-CoV-2 infection in the airways belong to multiple well-known pro-inflammatory pathways (Fig 3, S2 Table). Further data interpretation led us to classify disease-related proteins in two distinct sets of highly represented proinflammatory mediators and cytokines termed *Alarmins* and *Cytokine storm*, respectively (S1 Table). Alarmins represent a family of immunomodulatory proteins acting as damage-associated molecular patterns provided by injured stromal cells to recruit and activate various innate immune cells such as monocytes, macrophages, innate lymphoid cells as well as myeloid dendritic cells. Multiple proteins belonging to this family (*i.e*. defensins, HMGB1, IL1α, IL25, IL33, TSLP, S100A4, S100A7, S100A8, S100A9, S100A12, S100B, S100P) as well as their receptors such as IL1R1, RAGE, ST2 were predicted by our model to be involved in the evolution towards severe lung inflammation in COVID-19.

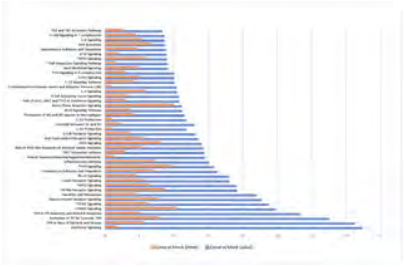

Fig 3. Pathway enrichment analysis from disease signatures (COVID-19 versus mock) in epithelial cell lines infected by SARS-CoV-2.
The top 40 most significantly dysregulated immunological canonical pathways in either the Calu-3 (yellow) and NHBE (brown) infected cell lines are represented in a radar plot according to -log (*p*-value). Pathway enrichment *z*-scores, based on fold change direction, represent predicted up-regulation (green dots) or down-regulation (blue dots) for positive or negative values, respectively.
https://doi.org/10.1371/journal.pone.0254374.g003

Our study also draws attention on disease-related proteins linked to the cytokine storm occurring in severe forms of COVID-19. The latter includes proinflammatory cytokines produced by activated myeloid cells such as IL1β, IL6 and TNFα directly involved as a cause of the CRS observed in COVID-19 [18, 35, 36]. Other potential targets associated with the cytokine storm include various cytokines (*e.g*. IL1β, IFNγ, IL2, IL12, IL15, IL17, IL23, IL32), chemokines (*e.g*. CCL5, CCL20, CXCL5, CXCL10, CXCL11), as well as selected proinflammatory factors (*e.g*. JAK1, JAK2, C5a) (S1 Table) [19, 20, 26–28, 36, 48–50].

Mapping into the interactome and identification of drug candidates for repurposing

COVID-19 disease-related proteins were mapped in parallel with known drug targets into the human complete interactome made of 15894 proteins (including 951 known drug targets) and 213861 interactions (Fig 4). From this, 3092 drugs were ranked according to computational proximity of their targets to each of the alarmins and cytokine storm sets by using a network-based method (S3 Table). Both COVID-19-related proteins as well as some functionally-related proteins in the interactome (such as the NR3C1 glucocorticoid receptor or receptors for reproductive steroids) were identified as candidate therapeutic targets.

Fig 4. Druggable interactome of proteins contributing to lung inflammation in COVID-19.
Extraction of the interactome encompassing proteins predicted to contribute to COVID-19 evolution towards a cytokine storm. Following SARS-CoV-2 infection of lung tissues and ensuing activation of innate and adaptive immune cells, different categories of proteins represent potential therapeutic targets to prevent or slow down lung inflammation associated with severe COVID-19. The latter include *Alarmins*, as well as cytokines, chemokines and selected proinflammatory factors associated with the *Cytokine storm*. For clarity, this figure only displays the disease related proteins (Alarmins & Cytokine storm) identified in our model, our top ranking repurposed drugs as well as some functional partners. The latter represent additional proteins needed in order to form a minimal principal component graph.
https://doi.org/10.1371/journal.pone.0254374.g004

Table 1 provides a list of selected targets as well as drugs interacting with those targets predicted to be of interest in severe COVID-19. Specifically, several high-ranking drugs were identified to treat severe COVID-19, such as anti-IL1β, anti-IL6 and IL6R or anti-TNFα antibodies. Our model supports as well the interest of corticosteroids such as dexamethasone, broadly used currently to treat severe COVID-19 [2]. Other high-ranking candidates for repurposing identified in our study are JAK2 inhibitors, with drugs not yet approved such as momelotinib or gandotinib previously shown by structure-based virtual screening to interact with ACE2 and the SARS-CoV-2 main protease, but also baricitinib, as well as other JAK1/JAK2 inhibitors currently being evaluated in COVID-19 patients (Table 1). Interestingly, some network computing approaches aiming to repurpose drugs inhibiting cell infection by SARS-CoV-2 also concluded to the interest of blocking antibodies against IL1β, IL6 and TNFα as well as JAK inhibitors in treating COVID-19 patients, in agreement with the present study [15, 16]. In addition, we also identify several reproductive steroids (estrogens and progesterone) as interesting candidates for treating COVID-19 patients.

Table 1. Overview of main therapeutic targets and clinical-stage candidate drugs for repurposing in COVID-19- related lung inflammation.
https://doi.org/10.1371/journal.pone.0254374.t001

Whereas the previous targets and some of the drugs directed to them could be expected from the current state of knowledge, our modeling study provided as well interesting hypotheses regarding other therapeutic options receiving less attention as of today. For example, drugs interacting with alarmins were also strongly suggested to be useful in COVID-19. To our knowledge, only three clinical studies have been initiated in COVID-19 with anti-alarmins, despite the availability of multiple additional drug candidates in

this class (Table 1). Noteworthy, since Alarmins of the S100 family activate Toll-like receptors such as TLR2 and TLR4, a therapeutic option might be to target specific TLRs downstream of alarmins. Indeed, several TLR-antagonists are currently undergoing clinical evaluation in order to restore immune-homeostasis in patients with COVID-19 [51].

Similarly, anti-IL17 antibodies rank very high in our repurposing analysis, suggesting that inhibitory drugs directed to this well-known pro-inflammatory cytokine as well as the functionally related IL23 cytokine or their receptors should be further investigated in COVID-19, with only one ongoing clinical trial in COVID-19 as of today [52]. In addition, the C5 complement inhibitor eculizumab is also predicted to represent an interesting treatment option, in agreement with recent evidence that the C5a-C5aR axis contributes to severe lung inflammation in COVID-19 patients [53]. As a strong chemoattractant, C5a provides in parallel to alarmins a link between innate and adaptive immune responses during severe COVID-19.

The thrombopoietin receptor appears as well to be a valid therapeutic target for agonists in light of the high incidence of thrombocytopenia associated with COVID-19 infection [54]. Rather unexpectedly, Topoisomerase 1 inhibitors, currently used as cytotoxic drugs in oncology, were also identified as of potential interest in COVID-19, with as of today only preclinical evidence that they can inhibit SARS-CoV-2 inflammation and death in animal models [55].

Supportive Cmap-based for drug repurposing

Given the rather limited set of transcriptomics data available and the small Cmap coverage for repurposable drugs (*i.e.* only 17% of molecules in our drug database, with none of the biologics), results were taken as supportive in the present study. Among the top network-based drugs proposed for repurposing, only 2 corticosteroids (betamethasone and hydrocortisone) were confirmed to elicit a reversed gene expression profile (Cmap score < -0.3) when compared to the disease gene expression state.

Conclusion

This study was designed to identify existing drugs which could be repurposed in a short time frame as a treatment for severe forms of COVID-19. We reasoned that such drugs should target those molecular pathways involved in the transition from mild lung inflammation caused by viral infection up to the cytokine storm associated with advanced stages of the disease (Fig 2, central and right lower panels). To this aim, using multiple sources of molecular profiling data from the literature relevant to distinguish mild from severe forms of the disease at the level of tissues and immune cells, we established a model of lung inflammation associated with COVID-19 in the form of an interactome of disease-related proteins. Combined with pharmacological knowledge of drug targets, this interactome allowed us to identify existing compounds which could be made available to patients in a short time frame.

Our network computational analyses identified several candidate therapeutic targets and corresponding drugs to repurpose which were confirmatory of existing knowledge (Table 1). This includes for example therapeutic antibodies interfering with either IL1β, IL6, TNFα or their receptors directly contributing to the CRS associated with severe COVID-19. Various inhibitory antibodies directed to these targets have already been evaluated in COVID-19 patients, such as anti-IL1® (canakinumab), anti-IL6R (tocilizumab, sarilumab) or anti-TNFα (infliximab, adalimumab) antibodies [4, 56]. Overall, these drugs yielded conflicting efficacy results, likely explained by evidence that such anti-cytokine treatments are rather effective if administered to patients before they develop advanced COVID-19 [57]. Nonetheless, a recent study evaluating the anti-IL6R antibodies tocilizumab and sarilumab demonstrated some improvement in survival when treating critically ill COVID-19 patients, even more so when these drugs were associated with dexamethasone [4, 5, 58]. Corticosteroids, are also predicted by the present study to be useful in severe COVID-19, in agreement with positive results previously obtained in multiple randomized clinical trials, eventually leading to a broad use of dexamethasone as a treatment for severe COVID-19 [2, 59]. JAK1 and JAK2 inhibitors came out also as interesting candidates for repurposing, with several inhibitors being actively tested in COVID-19 patients [60]. In this therapeutic class, the JAK1/JAK2 inhibitor baricitinib is currently raising most of the interest in light of recent evidence that it interferes with virus entry mediated by clathrin-associated endocytosis (Table 1) [61]. We also identified drugs interfering with reproductive steroids or their receptors as valid candidates for repurposing. This observation makes sense in light of the strong bias towards males among patients with severe COVID-19, perhaps explained in part by the upregulation by androgens of the expression of the SARS CoV-2 receptor [62]. In contrast estrogens and progesterone are rather considered to be protective in light of their anti-inflammatory properties as well as their capacity to promote proliferation and repair of respiratory epithelial cells [63]. On this basis, treatment with estrogens are being considered in patients with mild COVID-19 (Table 1).

Perhaps more interestingly, our repurposing study sheds light on other therapeutic classes which as of today receive insufficient attention as potential treatments for severe COVID-19. We predict that inhibitors of the well-known IL17 and IL23 proinflammatory cytokines (or their receptors) could be useful in COVID-19, with to our knowledge a single clinical trial evaluating as of today the anti-IL17 antibody secukinumab in COVID-19 [52]. Multiple monoclonal antibodies blocking those cytokines have been registered as treatments for other inflammatory diseases, which thus could be promptly repurposed in COVID-19 (Table 1). Similarly, the C5 complement inhibitor eculizumab was also identified to represent a valid therapeutic option, in agreement with recent evidence that the C5a-C5aR axis promotes severe lung inflammation in COVID-19 patients by mediating recruitment and activation of pro-inflammatory myeloid cells [53, 64]. Only proof of concept studies have been conducted so far in human with eculizumab, suggesting that this antibody may provide some benefit in severe COVID-19 [65, 66], with a confirmatory trial ongoing in a larger cohort of patients. Noteworthy, another clinical study has been recently initiated to evaluate as well in this indication the anti C5a receptor antibody avdoralimab (Table 1). Also, approaches combining JAK1/2 inhibitors with blockade of C5a with eculizumab are being considered as a treatment of severe pulmonary damage in COVID-19 patients [67]. Moreover, drugs such as romiplostim acting as an agonist for the thrombopoietin receptor are also predicted to be useful to treat COVID-19-associated thrombocytopenia, in agreement with a recent case study documenting platelet recovery following treatment with this drug of a COVID-19 pediatric patient [54].

The most significant outcome of our repurposing study is the prediction that several members of the alarmin family such as defensins, HMBG1, IL1α, IL25, IL33, TSLP, S100A4, S100A7, S100A8, S100A9, S100A12, S100B, S100P likely contribute to lung inflammation during COVID-19 (Fig 4) [68–70]. The role of each individual alarmin in this regard remains to be investigated, with presumably some of them (*e.g.* IL25, TSLP) rather contributing to the initial recruitment of myeloid cells and innate lymphoid cells following epithelial or endothelial cell infection, whereas others (IL33, S100 members) are likely being involved in later stages of lung inflammation culminating in the cytokine storm. The later assumption is consistent with recent observations that some alarmins can stimulate the production of both IL1β, IL6 and TNFα as well as multiple other proinflammatory cytokines and chemokines [71]. Furthermore, blood levels of IL1α, calprotectin (a heterodimer made of S100A8 and S100A9), S100A12, S100B and HGBM1 appear to correlate with COVID-19 severity [72–76] (S1 Table). Also, IL33 has been recently proposed to play a broad role in the pathophysiology of COVID-19 pneumonia by dampening both the antiviral interferon response as well as regulatory T cells, while promoting thrombosis and activating pro-inflammatory type 2 innate lymphoid cells and γδ T cells [77]. To our knowledge, only few clinical studies are being conducted as of today in COVID-19 with a TSLP inhibitor or with blocking antibodies directed to receptors

for IL1α or IL33 (i.e. ST2), whereas multiple additional blocking monoclonal antibodies directed to IL25, IL33 or TSLP are well under clinical evaluation to treat severe forms of asthma or atopic dermatitis [62, 69]. Furthermore, various inhibitors of the S100 family of proteins currently in preclinical development may represent promising drug candidates for the future (Table 1). We thus recommend considering existing anti-alarmins therapies to treat severe COVID-19, most particularly in the context of the converging rationale from this computational study as well as recent wet-lab evidence that this important class of proteins conveying proinflammatory signals plays a critical role in the pathophysiology of severe COVID-19. Lastly, this first model of severe lung inflammation in COVID-19 should be updated as new data are generated to better distinguish at an early stage patients with a high risk of evolving towards severe lung inflammation from those who will only develop mild forms of the disease.

Supporting information

S1 Table. Candidate COVID-19 related disease genes.
https://doi.org/10.1371/journal.pone.0254374.s001
(PDF)

S2 Table. Pathways enrichment analysis.
https://doi.org/10.1371/journal.pone.0254374.s002
(PDF)

S3 Table. Drug repurposing.
https://doi.org/10.1371/journal.pone.0254374.s003
(XLSX)

Acknowledgments

The authors are thankful to Dorothée Piva for providing excellent secretarial assistance.

References

1. Callaway E. Could new COVID variants undermine vaccines? Labs scramble to find out. Nature. 2021;589: 177–178. pmid:33432212
View Article • PubMed/NCBI • Google Scholar

2. Group TRC. Dexamethasone in Hospitalized Patients with Covid-19—Preliminary Report. New England Journal of Medicine. 2020 [cited 18 Jan 2021].

3. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the Treatment of Covid-19—Final Report. N Engl J Med. 2020;383: 1813–1826. pmid:32445440
View Article • PubMed/NCBI • Google Scholar

4. Salvarani C, Dolci G, Massari M, Merlo DF, Cavuto S, Savoldi L, et al. Effect of Tocilizumab vs Standard Care on Clinical Worsening in Patients Hospitalized With COVID-19 Pneumonia: A Randomized Clinical Trial. JAMA Intern Med. 2021;181: 24–31. pmid:33080005
View Article • PubMed/NCBI • Google Scholar

5. Hermine O, Mariette X, Tharaux P-L, Resche-Rigon M, Porcher R, Ravaud P, et al. Effect of Tocilizumab vs Usual Care in Adults Hospitalized With COVID-19 and Moderate or Severe Pneumonia: A Randomized Clinical Trial. JAMA Intern Med. 2021;181: 32–40. pmid:33080017
View Article • PubMed/NCBI • Google Scholar

6. Repurposed Antiviral Drugs for Covid-19—Interim WHO Solidarity Trial Results. New England Journal of Medicine. 2021;384: 497–511. pmid:33264556
View Article • PubMed/NCBI • Google Scholar

7. Krammer F. SARS-CoV-2 vaccines in development. Nature. 2020;586: 516–527. pmid:32967006
View Article • PubMed/NCBI • Google Scholar

8. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov. 2020;6: 14. pmid:32194980
View Article • PubMed/NCBI • Google Scholar

9. Nabirotchkin S, Peluffo AE, Bouaziz J, Cohen D. Focusing on the Unfolded Protein Response and Autophagy Related Pathways to Reposition Common Approved Drugs against COVID-19. 2020 [cited 18 Jan 2021].
View Article • Google Scholar

10. Li X, Yu J, Zhang Z, Ren J, Peluffo AE, Zhang W, et al. Network Bioinformatics Analysis Provides Insight into Drug Repurposing for COVID-2019. 2020 [cited 18 Jan 2021].
View Article • Google Scholar

11. Ciliberto G, Cardone L. Boosting the arsenal against COVID-19 through computational drug repurposing. Drug Discovery Today. 2020;25. pmid:32304645
View Article • PubMed/NCBI • Google Scholar

12. Chowdhury KH, Chowdhury MR, Mahmud S, Tareq AM, Hanif NB, Banu N, et al. Drug Repurposing Approach against Novel Coronavirus Disease (COVID-19) through Virtual Screening Targeting SARS-CoV-2 Main Protease. Biology. 2021;10: 2. pmid:33374717
View Article • PubMed/NCBI • Google Scholar

13. Stebbing J, Phelan A, Griffin I, Tucker C, Oechsle O, Smith D, et al. COVID-19: combining antiviral and anti-inflammatory treatments. Lancet Infect Dis. 2020;20: 400–402. pmid:32113509
View Article • PubMed/NCBI • Google Scholar

14. Gysi DM, Valle ÍD, Zitnik M, Ameli A, Gan X, Varol O, et al. Network Medicine Framework for Identifying Drug Repurposing Opportunities for COVID-19. arXiv:200407229 [cs, q-bio, stat]. 2020 [cited 1 Jun 2021]. Available: http://arxiv.org/abs/2004.07229 pmid:32550253
View Article • PubMed/NCBI • Google Scholar

15. Fiscon G, Conte F, Farina L, Paci P. SAveRUNNER: A network-based algorithm for drug repurposing and its application to COVID-19. PLOS Computational Biology. 2021;17: e1008686. pmid:33544720
View Article • PubMed/NCBI • Google Scholar

16. Fiscon G, Paci P. SAveRUNNER: An R-based tool for drug repurposing. BMC Bioinformatics. 2021;22: 150. pmid:33757425
View Article • PubMed/NCBI • Google Scholar

17. Tzotzos SJ, Fischer B, Fischer H, Zeitlinger M. Incidence of ARDS and outcomes in hospitalized patients with COVID-19: a global literature survey. Critical Care. 2020;24: 516. pmid:32825837
View Article • PubMed/NCBI • Google Scholar

18. Vardhana SA, Wolchok JD. The many faces of the anti-COVID immune response. J Exp Med. 2020;217. pmid:32353870
View Article • PubMed/NCBI • Google Scholar

19. Moore JB, June CH. Cytokine release syndrome in severe COVID-19. Science. 2020;368: 473–474. pmid:32303591
View Article • PubMed/NCBI • Google Scholar

20. de la Rica R, Borges M, Gonzalez-Freire M. COVID-19: In the Eye of the Cytokine Storm. Front Immunol. 2020;11. pmid:33072097
View Article • PubMed/NCBI • Google Scholar

21. Guney E, Menche J, Vidal M, Barábasi A-L. Network-based in silico drug efficacy screening. Nature Communications. 2016;7: 10331. pmid:26831545
View Article • PubMed/NCBI • Google Scholar

22. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabási A-L, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. Nat Commun. 2018;9: 2691. pmid:30002366
View Article • PubMed/NCBI • Google Scholar

23. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Møller R, Panis M, Sachs D, et al. SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. bioRxiv. 2020; 2020.03.24.004655.
View Article • Google Scholar

24. Ackermann M, Verleden SE, Kuehnel M, Haverich A, Welte T, Laenger F, et al. Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. New England Journal of Medicine. 2020;383: 120–128. pmid:32437596
View Article • PubMed/NCBI • Google Scholar

25. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nature Medicine. 2020;26: 842–844. pmid:32398875
View Article • PubMed/NCBI • Google Scholar

26. Laing AG, Lorenc A, del Molino del Barrio I, Das A, Fish M, Monin L, et al. A dynamic COVID-19 immune signature includes associations with poor prognosis. Nature Medicine. 2020;26: 1623–1635. pmid:32807934
View Article • PubMed/NCBI • Google Scholar

27. Hadjadj J, Yatim N, Barnabei L, Corneau A, Boussier J, Péré H, et al. Impaired type I interferon activity and exacerbated inflammatory responses in severe Covid-19 patients. medRxiv. 2020; 2020.04.19.20068015. pmid:32661059
View Article • PubMed/NCBI • Google Scholar

28. Ng LFP, Hibberd ML, Ooi E-E, Tang K-F, Neo S-Y, Tan J, et al. A human in vitro model system for investigating genome-wide host responses to SARS coronavirus infection. BMC Infect Dis. 2004;4: 34. pmid:15357874
View Article • PubMed/NCBI • Google Scholar

29. Brodin P. Immune determinants of COVID-19 disease presentation and severity. Nature Medicine. 2021;27: 28–33. pmid:33442016
View Article • PubMed/NCBI • Google Scholar

30. Wen W, Su W, Tang H, Le W, Zhang X, Zheng Y, et al. Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. Cell Discovery. 2020;6: 1–18. pmid:32377375
View Article • PubMed/NCBI • Google Scholar

31. Burke H, Freeman A, Cellura DC, Stuart BL, Brendish NJ, Poole S, et al. Inflammatory phenotyping predicts clinical outcome in COVID-19. Respiratory Research. 2020;21: 245. pmid:32962703
View Article • PubMed/NCBI • Google Scholar

32. Wu M, Chen Y, Xia H, Wang C, Tan CY, Cai X, et al. Transcriptional and proteomic insights into the host response in fatal COVID-19 cases. PNAS. 2020;117: 28336–28343. pmid:33082228
View Article • PubMed/NCBI • Google Scholar

33. Combes AJ, Courau T, Kuhn NF, Hu KH, Ray A, Chen WS, et al. Global absence and targeting of protective immune states in severe COVID-19. Nature. 2021; 1–10. pmid:33494096
View Article • PubMed/NCBI • Google Scholar

34. Arunachalam PS, Wimmers F, Mok CKP, Perera RAPM, Scott M, Hagan T, et al. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science. 2020;369: 1210–1220. pmid:32788292
View Article • PubMed/NCBI • Google Scholar

35. Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nature Reviews Immunology. 2020;20: 355–362. pmid:32376901
View Article • PubMed/NCBI • Google Scholar

36. Vabret N, Britton GJ, Gruber C, Hegde S, Kim J, Kuksin M, et al. Immunology of COVID-19: Current State of the Science. Immunity. 2020;52: 910–941. pmid:32505227
View Article • PubMed/NCBI • Google Scholar

37. Choreño-Parra JA, Jiménez-Álvarez LA, Cruz-Lagunas A, Rodríguez-Reyna TS, Ramírez-Martínez G, Sandoval-Vega M, et al. Clinical and immunological factors that distinguish COVID-19 from pandemic influenza A(H1N1). medRxiv. 2020; 2020.08.10.20170761.
View Article • Google Scholar

38. Atamas SP, Chapoval SP, Keegan AD. Cytokines in chronic respiratory diseases. F1000 Biol Rep. 2013;5. pmid:23413371
View Article • PubMed/NCBI • Google Scholar

39. Wang M, Withers JB, Ricchiuto P, Voitalov I, McAnally M, Sanchez HN, et al. A systems-based method to repurpose marketed therapeutics for antiviral use: a SARS-CoV-2 case study. Life Science Alliance. 2021;4. pmid:33593923
View Article • PubMed/NCBI • Google Scholar

40. Stolfi P, Manni L, Soligo M, Vergni D, Tieri P. Designing a Network Proximity-Based Drug Repurposing Strategy for COVID-19. Front Cell Dev Biol. 2020;8. pmid:33123533
View Article • PubMed/NCBI • Google Scholar

41. Cowen L, Ideker T, Raphael BJ, Sharan R. Network propagation: a universal amplifier of genetic associations. Nature Reviews Genetics. 2017;18: 551–562. pmid:28607512
View Article • PubMed/NCBI • Google Scholar

42. Song J-S, Wang R-S, Leopold JA, Loscalzo J. Network determinants of cardiovascular calcification and repositioned drug treatments. FASEB J. 2020;34: 11087–11100. pmid:32638415
View Article • PubMed/NCBI • Google Scholar

43. Chen X, Ji Z-L, Chen Y. TTD: Therapeutic Target Database. Nucleic acids research. 2002;30: 412–5. pmid:11752352
View Article • PubMed/NCBI • Google Scholar

44. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res. 2008;36: D901–906. pmid:18048412
View Article • PubMed/NCBI • Google Scholar

45. Cheng F, Kovács IA, Barabási A-L. Network-based prediction of drug combinations. Nature Communications. 2019;10: 1197. pmid:30867426
View Article • PubMed/NCBI • Google Scholar

46. Cao M, Zhang H, Park J, Daniels NM, Crovella ME, Cowen LJ, et al. Going the Distance for Protein Function Prediction: A New Distance Metric for Protein Interaction Networks. PLOS ONE. 2013;8: e76339. pmid:24194834
View Article • PubMed/NCBI • Google Scholar

47. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313: 1929–1935. pmid:17008526
View Article • PubMed/NCBI • Google Scholar

48. Sokulsky LA, Garcia-Netto K, Nguyen TH, Girkin JLN, Collison A, Mattes J, et al. A Critical Role for the CXCL3/CXCL5/CXCR2 Neutrophilic Chemotactic Axis in the Regulation of Type 2 Responses in a Model of Rhinoviral-Induced Asthma Exacerbation. The Journal of Immunology. 2020;205: 2468–2478. pmid:32948685
View Article • PubMed/NCBI • Google Scholar

49. Ye Q, Wang B, Mao J. The pathogenesis and treatment of the `Cytokine Storm' in COVID-19. J Infect. 2020;80: 607–613. pmid:32283152
View Article • PubMed/NCBI • Google Scholar

50. Leisman DE, Ronner L, Pinotti R, Taylor MD, Sinha P, Calfee CS, et al. Cytokine elevation in severe and critical COVID-19: a rapid systematic review, meta-analysis, and comparison with other inflammatory syndromes. The Lancet Respiratory Medicine. 2020;8: 1233–1244. pmid:33075298
View Article • PubMed/NCBI • Google Scholar

51. Information NC for B, Pike USNL of M 8600 R, MD B, Usa 20894. National Center for Biotechnology Information. [cited 19 May 2021]. https://www.ncbi.nlm.nih.gov/

52. Pacha O, Sallman MA, Evans SE. COVID-19: a case for inhibiting IL-17? Nature reviews Immunology. 2020;20: 345–346. pmid:32358580
View Article • PubMed/NCBI • Google Scholar

53. Carvelli J, Demaria O, Vély F, Batista L, Chouaki Benmansour N, Fares J, et al. Association of COVID-19 inflammation with activation of the C5a–C5aR1 axis. Nature. 2020;588: 146–150. pmid:32726800
View Article • PubMed/NCBI • Google Scholar

54. Schneider CW, Penney SW, Helfrich AM, Hartman KR, Lieuw K. A Novel Use of Romiplostim for SARS-CoV-2–induced Thrombocytopenia. Journal of Pediatric Hematology/Oncology. 2021;Publish Ahead of Print. pmid:33003146
View Article • PubMed/NCBI • Google Scholar

55. Ho JSY, Mok BW-Y, Campisi L, Jordan T, Yildiz S, Parameswaran S, et al. Topoisomerase 1 inhibition therapy protects against SARS-CoV-2-induced inflammation and death in animal models. bioRxiv. 2020; 2020.12.01.404483. pmid:33299999
View Article • PubMed/NCBI • Google Scholar

56. Robinson PC, Liew DFL, Liew JW, Monaco C, Richards D, Shivakumar S, et al. The Potential for Repurposing Anti-TNF as a Therapy for the Treatment of COVID-19. Med (N Y). 2020;1: 90–102. pmid:33294881
View Article • PubMed/NCBI • Google Scholar

57. De Stefano L, Bobbio-Pallavicini F, Manzo A, Montecucco C, Bugatti S. A “Window of Therapeutic Opportunity” for Anti-Cytokine Therapy in Patients With Coronavirus Disease 2019. Front Immunol. 2020;11. pmid:33123149
View Article • PubMed/NCBI • Google Scholar

58. Della-Torre E, Campochiaro C, Cavalli G, De Luca G, Napolitano A, La Marca S, et al. Interleukin-6 blockade with sarilumab in severe COVID-19 pneumonia with systemic hyperinflammation: an open-label cohort study. Ann Rheum Dis. 2020;79: 1277–1285. pmid:32620597
View Article • PubMed/NCBI • Google Scholar

59. WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group, Sterne JAC, Murthy S, Diaz JV, Slutsky AS, Villar J, et al. Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis. JAMA. 2020;324: 1330–1341. pmid:32876694
View Article • PubMed/NCBI • Google Scholar

60. Luo W, Li Y-X, Jiang L-J, Chen Q, Wang T, Ye D-W. Targeting JAK-STAT Signaling to Control Cytokine Release Syndrome in COVID-19. Trends in Pharmacological Sciences. 2020;41: 531–543. pmid:32580895
View Article • PubMed/NCBI • Google Scholar

61. Seif F, Aazami H, Khoshmirsafa M, Kamali M, Mohsenzadegan M, Pornour M, et al. JAK Inhibition as a New Treatment Strategy for Patients with COVID-19. Int Arch Allergy Immunol. 2020;181: 467–475. pmid:32392562
View Article • PubMed/NCBI • Google Scholar

62. Fagone P, Ciurleo R, Lombardo SD, Iacobello C, Palermo CI, Shoenfeld Y, et al. Transcriptional landscape of SARS-CoV-2 infection dismantles pathogenic pathways activated by the virus, proposes unique sex-specific differences and predicts tailored therapeutic strategies. Autoimmunity Reviews. 2020;19: 102571. pmid:32376402
View Article • PubMed/NCBI • Google Scholar

63. Pinna G. Sex and COVID-19: A Protective Role for Reproductive Steroids. Trends in Endocrinology & Metabolism. 2021;32: 3–6. pmid:33229187
View Article • PubMed/NCBI • Google Scholar

64. Peffault de Latour R, Bergeron A, Lengline E, Dupont T, Marchal A, Galicier L, et al. Complement C5 inhibition in patients with COVID-19—a promising target? Haematologica. 2020;105: 2847–2850. pmid:33256385
View Article • PubMed/NCBI • Google Scholar

65. Annane D, Heming N, Grimaldi-Bensouda L, Frémeaux-Bacchi V, Vigan M, Roux A-L, et al. Eculizumab as an emergency treatment for adult patients with severe COVID-19 in the intensive care unit: A proof-of-concept study. EClinicalMedicine. 2020;28. pmid:33173853
View Article • PubMed/NCBI • Google Scholar

66. Diurno F, Numis FG, Porta G, Cirillo F, Maddaluno S, Ragozzino A, et al. Eculizumab treatment in patients with COVID-19: preliminary results from real life ASL Napoli 2 Nord experience. Eur Rev Med Pharmacol Sci. 2020;24: 4040–4047. pmid:32329881
View Article • PubMed/NCBI • Google Scholar

67. Giudice V, Pagliano P, Vatrella A, Masullo A, Poto S, Polverino BM, et al. Combination of Ruxolitinib and Eculizumab for Treatment of Severe SARS-CoV-2-Related Acute Respiratory Distress Syndrome: A Controlled Study. Front Pharmacol. 2020;11: 857. pmid:32581810
View Article • PubMed/NCBI • Google Scholar

68. Yalcin Kehribar D, Cihangiroglu M, Sehmen E, Avci B, Capraz A, Yildirim Bilgin A, et al. The receptor for advanced glycation end product (RAGE) pathway in COVID-19. Biomarkers. 2021; 1–5. pmid:33284049
View Article • PubMed/NCBI • Google Scholar

69. Roth A, Lütke S, Meinberger D, Hermes G, Sengle G, Koch M, et al. LL-37 fights SARS-CoV-2: The Vitamin D-Inducible Peptide LL-37 Inhibits Binding of SARS-CoV-2 Spike Protein to its Cellular Receptor Angiotensin Converting Enzyme 2 In Vitro. bioRxiv. 2020; 2020.12.02.408153.
View Article • Google Scholar

70. Idris MM, Banu S, Siva AB, Nagaraj R. Downregulation of Defensin genes in SARS-CoV-2 infection. medRxiv. 2020; 2020.09.21.20195537.
View Article • Google Scholar

71. Yang D, Han Z, Oppenheim JJ. ALARMINS AND IMMUNITY. Immunol Rev. 2017;280: 41–56. pmid:29027222
View Article • PubMed/NCBI • Google Scholar

72. Chen L, Long X, Xu Q, Tan J, Wang G, Cao Y, et al. Elevated serum levels of S100A8/A9 and HMGB1 at hospital admission are correlated with inferior clinical outcomes in COVID-19 patients. Cellular & Molecular Immunology. 2020;17: 992–994. pmid:32620787
View Article • PubMed/NCBI • Google Scholar

73. Silvin A, Chapuis N, Dunsmore G, Goubet A-G, Dubuisson A, Derosa L, et al. Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19. Cell. 2020;182: 1401–1418.e18. pmid:32810439
View Article • PubMed/NCBI • Google Scholar

74. Zuniga M, Gomes C, Carsons SE, Bender MT, Cotzia P, Miao QR, et al. Autoimmunity to the Lung Protective Phospholipid-Binding Protein Annexin A2 Predicts Mortality Among Hospitalized COVID-19 Patients. medRxiv. 2021; 2020.12.28.20248807.
View Article • Google Scholar

75. Aceti A, Margarucci LM, Scaramucci E, Orsini M, Salerno G, Di Sante G, et al. Serum S100B protein as a marker of severity in Covid-19 patients. Scientific Reports. 2020;10: 18665. pmid:33122776
View Article • PubMed/NCBI • Google Scholar

76. Zeng Z, Hong X-Y, Li Y, Chen W, Ye G, Li Y, et al. Serum-soluble ST2 as a novel biomarker reflecting inflammatory status and illness severity in patients with COVID-19. Biomarkers in Medicine. 2020;14: 1619–1629. pmid:33336592
View Article • PubMed/NCBI • Google Scholar

77. Zizzo G, Cohen PL. Imperfect storm: is interleukin-33 the Achilles heel of COVID-19? The Lancet Rheumatology. 2020;2: e779–e790. pmid:33073244
View Article • PubMed/NCBI • Google Scholar