

Mesure de la performance à une tâche de décision dynamique dans un micromonde simulé par ordinateur

BENOÎT BÉCHARD, ÉCOLE DE PSYCHOLOGIE
UNIVERSITÉ LAVAL
benoit.bechar.d.1@ulaval.ca

FRÉDÉRIC MORNEAU-GUÉRIN, DÉPARTEMENT ÉDUCATION
TÉLUQ
frederic.morneau-guerin@teluq.ca

Résumé

L'étude de la prise de décision dans les environnements complexes du monde réel comporte certains défis. Il est notamment difficile d'étudier l'effet de la complexité sur les processus cognitifs impliqués dans la prise de décision ; une approche par essais et erreurs est rarement souhaitable dans le monde réel, puisque de mauvaises décisions peuvent avoir des conséquences désastreuses. La simulation par micromonde s'est avérée efficace pour étudier expérimentalement la prise de décision dynamique dans des environnements capturant les propriétés de la complexité du monde réel. Toutefois, afin de saisir pleinement l'impact de la complexité sur la décision humaine, il est essentiel d'évaluer la performance en tenant compte de la *polytely*, la caractéristique de la complexité qui se rapporte précisément à l'exécution de la tâche, et qui implique la nécessité pour le décideur de concilier des objectifs conflictuels. À travers une approche d'analytique du jeu basée sur *Democracy 3*, nous présentons une méthode visant à dériver un indicateur mathématique de la performance, lequel permet de saisir l'importance de la complexité de l'environnement dans la réalisation d'une tâche de prise de décision dynamique.

Mots clés : Mesure de la performance ; psychologie mathématique ; décision dynamique ; psychologie cognitive

Classification mathématique par matières (MSC2020) : 00A06, 91C05, 91E10, 91E45

La complexité dynamique est la principale caractéristique du comportement contre-intuitif des systèmes complexes [48]. Les environnements de décision du monde réel sont complexes en ce qu'ils comprennent un grand nombre de variables mutuellement dépendantes (connectivité), qui évoluent de manière imprévisible et non linéaire dans le temps et l'espace (dynamisme). Les décideurs agissant dans ces environnements sont soumis à des niveaux variables d'opacité, en ce

que les informations pertinentes pour la décision peuvent n'être que partiellement accessibles dans l'accomplissement d'une tâche visant à concilier des objectifs entrant parfois en contradiction (*polytely*) [20, 40]. Comprendre le comportement contre-intuitif des environnements du monde réel est nécessaire dans tous les domaines de l'activité humaine caractérisés par une certaine forme de complexité dynamique, qu'il s'agisse de la médecine d'urgence, de la planification financière, de la gestion de crises, ou de l'élaboration de politiques publiques [18, 28, 36].

Dès qu'un environnement de décision met en relation plus de trois variables – une relation de complexité quaternaire (voir [26]) – la recherche mentionne que la cognition humaine est sévèrement limitée dans sa capacité à comprendre et prévoir son comportement. Malgré leur faible niveau de complexité, ces systèmes créent néanmoins un état de surcharge cognitive chez le décideur [30], l'incitant à adopter une approche « réductionniste » du traitement de l'information pour résoudre les problèmes. Lorsqu'ils sont confrontés à des problèmes complexes et mal définis (voir [18]), les individus peuvent se concentrer sur des objectifs à court terme (au lieu d'objectifs à long terme) ([25, 31]), considérer ces objectifs comme indépendants et dissociés, rechercher une cause unique pour expliquer les situations problématiques, anticiper l'avenir comme une projection linéaire, tout en minimisant les effets secondaires potentiels et les tendances exponentielles [10, 12]. Fréquemment associée à une vision en tunnel [38] découlant d'une formulation partielle de la problématique à traiter et d'une évaluation insuffisante de la situation, l'adoption d'une telle approche de résolution de problèmes compromet la décision. Diehl et Sterman [7] ont notamment démontré que l'application d'une règle simple de « ne rien faire » peut parfois s'avérer plus efficace que l'intervention humaine pour contrôler des systèmes complexes (voir également [33, 42]).

La résolution de problèmes complexes (*complex problem solving*, ou RPC) se concentre notamment sur l'identification des déterminants de la performance dans des tâches de prise de décision prenant place dans des environnements soumis à la complexité dynamique. La recherche sur la RPC utilise des environnements de simulation interactifs, souvent appelés micromondes [23], pour reproduire les propriétés des environnements complexes qui occupent les décideurs dans le monde réel. Les micromondes sont des univers de jeu exploratoires, ouverts et polyvalents, car ils permettent de configurer des tâches de décision dynamique pour la recherche en cognition. Dans une perspective de cognition dynamique, où les mécanismes de la pensée humaine émergent de

manière adaptative et interactive au fur et à mesure de l'exécution d'une tâche, les chercheurs en RPC ont utilisé la simulation par micromonde pour approfondir notre compréhension de la relation entre l'environnement et les processus cognitifs dans la détermination du comportement de décision face à la complexité (voir [11, 16]).

La complexité d'un environnement de simulation peut varier et, de ce fait, la difficulté pour un décideur humain d'en contrôler l'évolution. Comment alors garantir qu'une mesure de la performance considère adéquatement la complexité dans l'exécution d'une tâche de décision dynamique? Dans la recherche en RPC, les chercheurs ne sont pas parvenus à un consensus quant à la meilleure manière de mesurer le niveau de complexité d'un environnement de simulation, ni non plus son impact sur le développement des compétences de résolution de problèmes (p. ex., [24, 19]). Rigas *et coll.* [45] souligne notamment que le rôle de l'approche par micromonde est de capturer les propriétés de la complexité que l'on rencontre dans la vie réelle, et que supprimer l'un ou l'autre de ces attributs dans la mesure de la RPC revient à tenter de mesurer un concept extérieur à la complexité. Un aspect important de l'évaluation de la performance consiste à s'assurer que non seulement l'environnement de simulation capture les propriétés de la complexité, mais également que la mesure en elle-même contribue à cette capture, de manière à refléter la difficulté que rencontre un décideur lors de la gestion d'objectifs contradictoires dans le monde réel. Ainsi, un indicateur de performance doit considérer le niveau de difficulté (complexité) du micromonde utilisé. Dans cette étude, nous proposons une mesure ouverte de la performance à une tâche de RPC qui considère le niveau de complexité de l'environnement de simulation. L'indicateur proposé tient compte de la difficulté de concilier des objectifs contradictoires au sein d'un environnement complexe. En testant l'indicateur au moyen d'un échantillon aléatoire généré en fonction des paramètres du jeu de simulation *Democracy 3* utilisé comme micromonde de recherche, nous contribuons à démontrer que la méthode proposée peut être utile pour évaluer dans quelle mesure un décideur soumis à la *polytely* répond aux objectifs de décision dans un environnement complexe.

1 Objectif de l'étude

L'objectif de l'étude est d'évaluer si un indice qui capture la difficulté de concilier des objectifs contradictoires dans un environnement complexe peut servir de mesure de la performance transférable à d'autres contextes de prise de décision dans le monde réel. Parmi les propriétés de la complexité identifiées et reconnues dans la littérature sur la RPC, la *polytely* est la caractéristique qui se rapporte spécifiquement à l'exécution d'une tâche, ainsi qu'à l'effort cognitif réalisé par un décideur pour atteindre ses objectifs dans un environnement complexe. Grâce à un processus de modélisation formelle, nous mettons en évidence les avantages potentiels pour l'opérationnalisation de la RPC de l'utilisation d'un micromonde en combinaison avec un indice ouvert qui saisit la complexité dynamique.

Nous proposons d'abord une brève description de l'environnement de simulation utilisé, puis nous détaillons une procédure mathématique permettant d'élaborer un indicateur multidimensionnel de la performance à une tâche de RPC. L'indicateur est développé afin de tenir compte de la nature conflictuelle des objectifs de décision dans le monde réel, et permet d'identifier dans quelle mesure le comportement du décideur a répondu aux besoins de la tâche. Un échantillon de données est ensuite généré aléatoirement pour tester la robustesse de l'indicateur comme mesure de la complexité dynamique, et vérifier si ce dernier se comporte de manière cohérente pour permettre des comparaisons inter-sujets du comportement décisionnel, indépendamment de la nature ou de l'échelle de leurs réponses [15]. À travers un système de notation facilitant la comparaison dans un même espace solution (à ce sujet, voir [31]), l'indicateur mesure la performance à une tâche de RPC exigeant du décideur qu'il concilie des objectifs contradictoires en contrôlant un système soumis à la complexité dynamique. En outre, la procédure permet de souligner la pertinence de l'indicateur pour mesurer la performance dans d'autres contextes qui s'appuient sur la simulation par micromonde [28].

2 *Democracy 3* comme micromonde de recherche

Les micromondes sont notamment utilisés pour étudier le contrôle et la résolution de problèmes complexes [11] et la prise de décision [22] dans le cadre d'expériences contrôlées en laboratoire (p.

ex., [3]). Les micromondes simulent les caractéristiques de tâches réelles qui se concentrent sur des scénarios ouverts. Ils offrent un récit convaincant des nombreux défis cognitifs auxquels un individu est confronté lorsqu'il doit gérer des problèmes complexes (p. ex., pénurie de ressources, pression temporelle, délais imprévus, développements exponentiels). Dans une tâche de RPC par micromonde, les participants doivent traiter un certain nombre de variables d'entrée, observer et expérimenter les changements dans le système par le biais d'un retour d'information (en fonction d'un certain nombre de variables de sortie), et ainsi, mettre à jour leurs connaissances et leur représentation mentale de l'espace problème. En simulant un environnement cognitivement exigeant (similaire à ceux rencontrés dans le monde réel), les micromondes permettent aux chercheurs de collecter des données comportementales sur le traitement de l'information et la prise de décision pour mieux comprendre les limites de la cognition humaine face à la complexité dynamique [35].

Les micromondes s'appuient sur la dynamique systémique (*system dynamics*) pour modéliser la complexité. Par le biais de diagrammes d'influence (cause-effet), cette méthodologie permet de représenter un grand nombre de variables mutuellement dépendantes et interconnectées pour illustrer et décrire le comportement non linéaire des environnements complexes [21]. S'appuyant sur l'exploration, l'essai-erreur, et les boucles de rétroaction, les modèles de dynamique systémique fournissent des informations au décideur sous la forme de rétroactions plus ou moins détaillées. Comme dans la vie réelle, il n'existe pas de « bonnes » ou de « mauvaises » réponses prédéterminées ; les décisions sont prises dans un environnement qui change en fonction de l'intervention du décideur, mais aussi de manière autonome et imprévisible au fil du temps. Il est particulièrement difficile pour le décideur aux prises avec une tâche de RPC de prédire l'ampleur des conséquences (positives ou négatives) résultant de ses décisions. En capturant les propriétés de la complexité du monde réel grâce à la dynamique systémique, les micromondes jouent un rôle essentiel dans la recherche sur la RPC en psychologie expérimentale (voir [22, 23]).

Plusieurs jeux de simulation ont été développés pour répondre à des objectifs autres que strictement vidéoludiques¹, que ce soit sur le plan de l'éducation et de l'apprentissage (p. ex., [46]), promouvoir l'engagement social (p. ex., [1]), ou encore accompagner les décideurs

1. Stokes [49] définit les *jeux sérieux* comme des jeux dont l'objectif principal dépasse le simple divertissement pour s'orienter vers l'éducation, la sensibilisation, ou encore la formation.

dans la sélection du personnel dans les organisations (p. ex., [39, 50]). D'autres simulations, initialement développées à des fins purement ludiques, peuvent être mises au service de la recherche (p. ex., *Democracy 3*, [4]; *Cities : Skylines*, [2]; *floodsim*, [44]). Néanmoins, il convient de noter que toutes les simulations ne peuvent pas être utilisées en tant que micromondes de recherche. En ce qui concerne la recherche en RPC, il est nécessaire d'observer le comportement des participants et de convertir les données en informations exploitables pour identifier les déterminants de la performance [35]. Pour ce faire, la plateforme utilisée doit adéquatement capturer les propriétés de la complexité dynamique et offrir un accès aux données brutes sur les relations fonctionnelles entre le comportement du décideur et l'environnement de simulation (voir [8]). Les simulations qui offrent la possibilité d'intégrer des mesures complémentaires (sans compromettre l'intégrité de la tâche) sont à privilégier, car elles permettent la triangulation des données, ainsi que leur validation [5]. Des exemples pertinents se retrouvent notamment dans des scénarios de simulation classiques, tels que ceux présentés dans les travaux du chercheur et psychologue allemand Dietrich Dörner comme « Tanaland » et « Greenvale » [9, 14], ou encore « Lohhausen » [13]. D'autres exemples peuvent être observés sur des plateformes de simulation pour la recherche, tels que *CODEM* (voir [32]), ainsi que dans des simulations ludiques de prise de décision comme *Ecopolicy* [51], ou encore *Democracy* [43].

Dans le cadre de la présente étude, *Democracy 3* (D3) [43] a été utilisé comme micromonde de recherche. D3 est un simulateur politique ouvert qui capture les propriétés inhérentes à la complexité du monde réel. Dans D3, le joueur doit prendre plusieurs décisions afin d'atteindre les objectifs du jeu. Dans ce contexte, les nombreux éléments qui constituent l'environnement sont étroitement interdépendants, et l'espace problème évolue de manière autonome (c.-à-d., avec ou sans l'intervention du joueur). Chacune des décisions a un impact en temps réel sur le système (à cet effet, voir les travaux de [6, 22]). Le joueur est placé dans la position d'un chef d'état et sa tâche consiste à assurer la gestion des affaires publiques et politiques d'un pays du monde réel. Spécifiquement, le joueur doit prendre des décisions politiques pour accroître le soutien du public à son gouvernement, tout en assurant une gestion efficace des finances publiques – dans le but d'être réélu à la fin de la séance de jeu [27]. Grâce à un scénario de résolution de problèmes ouvert, D3 recrée des défis décisionnels dans sept différents domaines de la vie sociale, à savoir l'économie, les affaires étrangères, la loi et l'ordre, les services publics,

les impôts, les transports, et l'aide sociale. Le jeu s'appuie sur la dynamique systémique pour modéliser les propriétés de la complexité. L'interface principale comprend un grand nombre d'objets représentés par des icônes circulaires bleues (statistiques), blanches (politiques actives), rouges (situations problématiques), ou vertes (avantages technologiques) (voir Figure 1).

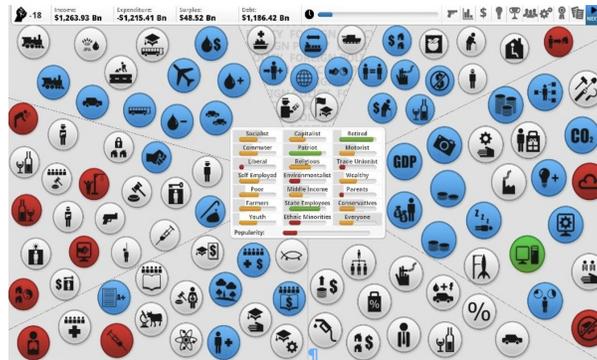


FIGURE 1 – Interface principale de D3 avec représentation iconique de la complexité politique.

Chaque objet est interconnecté par des relations d'effets représentées par des flèches rouges ou vertes qui se déplacent à une vitesse proportionnelle à la force de leur impact sur la situation cible (voir Figure 2). Chacun de ces objets évolue de manière dynamique dans le temps et dans l'espace en réaction aux décisions (ou en l'absence de décision) prises par le joueur. La dynamique systémique est relativement transparente (c.-à-d., bien que le joueur ne puisse pas connaître la nature exacte de l'effet d'une décision sur le système, une estimation de la popularité de chaque politique publique pouvant être mise en oeuvre est fournie, ainsi que le diagramme d'influence qui sous-tend les relations entre les différentes variables du jeu), et elle fournit un retour d'information au joueur sur la base de ses interactions avec le système.

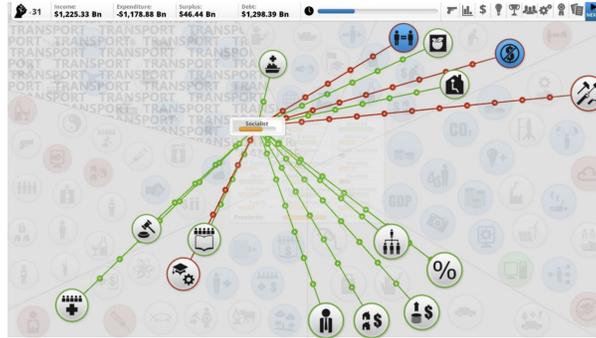


FIGURE 2 – Représentation de l’opacité partielle de la dynamique systémique de D3.

D3 a été conçu dans le but de permettre aux joueurs qui ne sont pas familiers avec la programmation de modifier le scénario d’origine en ajoutant de nouveaux éléments au jeu, tels que des situations, des dilemmes, ou encore de nouvelles politiques. Un utilisateur peut aussi supprimer des variables existantes, ou encore modifier la grandeur de l’impact des décisions prises par le joueur à des fins de recherche. La plateforme permet l’extraction de données brutes issues du comportement du joueur dans le jeu. D3 fournit également au joueur des indicateurs de suivi en temps réel, notamment en ce qui a trait à la popularité (p. ex., le pourcentage de la population qui semble favorable au gouvernement), et à la situation budgétaire. Ces deux indicateurs sont accessibles sous forme d’indices numériques pouvant être extraits des fichiers *.txt* ou *.csv* à l’aide d’un simple éditeur de texte. Ces caractéristiques font de D3 une plateforme polyvalente pour la recherche, en ce qu’elle permet la modification du scénario du jeu et des règles de logique qui organisent la dynamique systémique, afin de tester des hypothèses spécifiques sur la prise de décision.

3 Une mesure de la réalisation des objectifs

Nous souhaitons un indice qui puisse capturer la *polytely* (c.-à-d., la présence d’objectifs contradictoires dans la gestion et la résolution de problèmes complexes). La métrique de performance devait modéliser un comportement particulier – *ceteris paribus* – lorsque le pourcentage de votes reçus variait (p. ex., un pourcentage plus élevé conduit à une meilleure

performance), lorsque le budget variait (p. ex., un équilibre budgétaire conduit à une meilleure performance), et lorsque le résultat de l'élection variait (p. ex., la réélection conduit à une meilleure performance). Nous avons constaté que la métrique qui répondait le mieux à ces critères consistait en un produit de trois fonctions basées sur les indicateurs intégrés dans D3. Cette mesure visait à évaluer le degré de réalisation des différents objectifs en se basant sur la variation du pourcentage de votes reçus par le joueur à la fin du jeu, la situation budgétaire lors du dernier tour de jeu, ainsi que le résultat de l'élection – que le joueur soit réélu ou non. L'indicateur a été utilisé pour évaluer dans quelle mesure un participant répondait aux exigences de la tâche de RPC, à savoir prendre des décisions politiques pour accroître le soutien du public à l'égard du gouvernement, tout en gérant au mieux les finances public en vue d'une réélection. L'indicateur a servi à mesurer dans quelle mesure un participant était capable de satisfaire les exigences de la tâche de RPC, laquelle consistait à prendre des décisions politiques pour augmenter le soutien du public envers le gouvernement, tout en assurant une gestion efficace des finances publiques en vue de sa réélection. Une description détaillée de la méthode utilisée pour développer chaque fonction est présentée dans la section suivante.

Description de l'indicateur de performance

Notre objectif consistait à définir une fonction F qui attribue un score de performance unique — à savoir un nombre réel non négatif — à chaque triplet (x, y, z) , où

- x désigne le *pourcentage de votes reçus*, un nombre réel compris entre 0 (c'est-à-dire 0%) et 1 (c'est-à-dire 100%);
- y est le *montant final du budget de fonctionnement*, un nombre réel (un nombre négatif représente un déficit tandis qu'un nombre positif représente un excédent);
- z tient compte du *résultat de l'élection* et prend ses valeurs dans l'ensemble {réélu, battu};

afin que les performances puissent être ordonnées et comparées.

Compte tenu du contexte et des applications souhaitées, certains comportements spécifiques sont attendus de cette fonction $F(x, y, z)$ lorsque le vote varie (*ceteris paribus*), lorsque le budget varie (*ceteris paribus*) et lorsque les résultats de l'élection varient (*ceteris paribus*). Cela suggère que nous définissions d'abord trois fonctions, à savoir $V(x)$, $B(y)$ et $E(z)$, affichant les

comportements attendus, et que nous considérons ensuite l'indicateur de performance global de x , y et z comme l'agrégation multiplicative de $V(x)$, $B(y)$ et $E(z)$. Plus formellement, on définit

$$F : (0, 1) \times (-\infty, \infty) \times \{\text{réélu, battu}\} \longrightarrow [0, \infty)$$

comme la moyenne géométrique de $V(x)$, $B(y)$ et $E(z)$:

$$F(x, y, z) := \sqrt[3]{V(x)B(y)E(z)}. \quad (3.1)$$

Pourcentage de votes reçus

D'une part, comme il est plus facile de convaincre les électeurs indécis de voter pour soi que de rallier les électeurs formant la base électorale d'un autre parti, il semble raisonnable que la fonction traitant du pourcentage de votes reçus soit non seulement monotone croissante sur son domaine de définition, mais aussi strictement convexe (ce qui signifie qu'elle sera constamment plus abrupte à mesure que le pourcentage de votes reçus augmente).

D'autre part, compte tenu de la forme multiplicative de notre indicateur de performance global, il apparaît souhaitable que la fonction traitant du pourcentage de votes reçus tende vers 0 lorsque le pourcentage tend vers 0.

Parmi les nombreuses fonctions élémentaires qui ont le comportement souhaité, l'une d'entre elles — à savoir la *tangente hyperbolique inverse* — se distingue par le fait que son domaine de définition répond adéquatement à nos besoins.

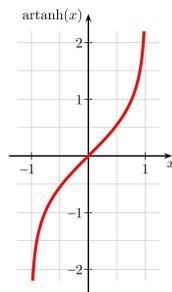


FIGURE 3 – Graphe de $\operatorname{artanh}(x)$, la tangente hyperbolique inverse.

On définit donc une fonction $V : [0, 1) \rightarrow [0, \infty)$ comme suit :

$$V(x) := c \times (x), \quad (3.2)$$

où $c > 0$ est un paramètre de mise à l'échelle dont la valeur reste à fixer.

Montant final du budget de fonctionnement

Notre fonction mesurant la performance par rapport à la situation budgétaire doit punir les dettes et récompenser les excédents budgétaires, l'ampleur de la punition ou de la récompense devant être amplifiée à mesure que le module de la dette ou de l'excédent augmente. Quant au fait de terminer le jeu avec un budget parfaitement équilibré, il ne doit être ni sanctionné ni récompensé.

Une fonction répondant à ces critères aurait donc une courbe en forme de « S ». Une telle fonction est appelée *sigmoïde*. Il existe une grande variété de fonctions sigmoïdes, l'une des plus simples et des plus connues est la *fonction logistique* $B : (-\infty, \infty) \rightarrow (0, \infty)$ définie par :

$$B(y) := \frac{L}{1 + \exp(-ky)}, \quad (3.3)$$

où

- $L > 0$ désigne la somme des valeurs de la fonction (c'est-à-dire l'asymptote horizontale supérieure) ;
- $k > 0$ représente le taux de croissance.

Pour des valeurs de y parcourant l'ensemble de la droite réelle on obtient la courbe sigmoïde ci-dessous, où le graphe de f tend vers L lorsque x s'approche de ∞ et tend vers zéro lorsque x s'approche de $-\infty$. Les deux asymptotes sont approchées par la courbe de façon symétrique.

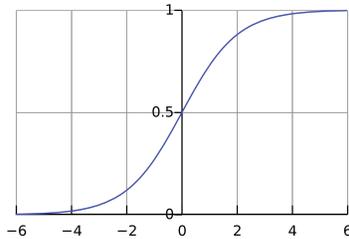


FIGURE 4 – Fonction logistique, où $L = 1$ et $k = 1$.

La valeur du supremum L et celle du paramètre de mise à l'échelle k influençant la pente de la courbe seront déterminées ultérieurement.

Résultat de l'élection

Une élection étant un phénomène dichotomique (c'est-à-dire qu'un participant est soit élu, soit battu), la fonction E permettant de prendre en compte le résultat de l'élection doit être binaire. Compte tenu du caractère multiplicatif de notre indicateur de performance globale, il apparaît pertinent de définir E comme retournant 1 en cas de réélection et une pénalité — un nombre positif inférieur à 1 — en cas de défaite du participant. Une réélection aura donc un effet neutre sur le score de performance, tandis qu'une défaite électorale aura pour effet de réduire la performance globale.

3.1 Paramètres de mise à l'échelle

Pourcentage de votes reçus

Compte tenu de la méthode de détermination de la valeur du paramètre de mise à l'échelle de la fonction traitant du pourcentage de votes reçus détaillée ci-dessus, la grande majorité des valeurs retournées par la fonction V lorsqu'on lui donne des données expérimentales seront dans l'intervalle $[0, 2]$.

Il semble souhaitable que les fonctions $V(x)$ et $B(y)$ soient du même ordre de grandeur, sinon il y a le risque que l'une de ces fonctions masque complètement l'effet de l'autre. Il semble donc approprié de fixer $L = 2$ dans (3.3). Par conséquent, si un participant à l'étude termine le jeu avec un budget parfaitement équilibré, la fonction B renverra la valeur 1. Tout montant de dette renverrait une valeur comprise entre 0 et 1, et tout excédent renverrait une valeur comprise entre 1 et 2.

Quant à la valeur du paramètre influençant l'inclinaison de la pente, elle doit être fixée de manière à ce que la valeur de $B(y)$ ne converge ni trop lentement ni trop rapidement vers ses

valeurs extrêmes. Comme la plupart des participants réalisent des performances sous-optimales, nous nous intéresserons principalement au comportement de $B(y)$ pour des valeurs situées sur la demi-droite inférieure.

Une façon simple de s'assurer que $B(y)$ ne domine pas fortement $V(x)$ et qu'il n'est pas non plus fortement dominé par lui consiste à définir

$$V(x_{\min}) = B(y_0),$$

où

- x_{\min} correspond au plus petit pourcentage de votes reçus par un participant dans notre échantillon ;
- y_0 désigne un montant final du budget de fonctionnement qui est le même nombre d'écart absolu médian par rapport au montant médian du budget de fonctionnement que x_{\min} est par rapport au pourcentage médian des votes reçus.

Résultat de l'élection

Comme nous l'avons vu plus haut, la fonction qui traite du résultat de l'élection est binaire. Son effet est imperceptible si le participant est réélu et elle inflige un malus si le participant est battu. La valeur du malus est inévitablement subjective. Afin de réduire autant que possible l'arbitraire, nous proposons une manière systématique de fixer sa valeur en utilisant l'ensemble des données.

Par exemple, la valeur du malus peut être fixée de manière à ce que l'indicateur de performance global d'un participant qui a réalisé une performance légèrement supérieure à la médiane (par ex. de $\frac{1}{4}$ MAD_ν par rapport à x et y) mais qui a été battu soit abaissé au même niveau que celui d'un participant qui a réalisé une performance légèrement inférieure à la médiane (par exemple de $\frac{1}{4}$ MAD_ν par rapport à x et y) et qui a néanmoins été réélu.

Un indicateur de performance global fonctionnel

Avec $V(x)$, $B(y)$ et $E(z)$ ainsi définis, nous sommes maintenant en mesure d'attribuer un indicateur de performance globale à tout triplet (x, y, z) comme suit :

$$F(x, y, z) := \sqrt[3]{V(x)B(y)E(z)}.$$

Bien qu'en théorie l'image de cette fonction soit la demi-droite $[0, \infty)$, en pratique, les valeurs attribuées aux triplets qui sont jugées plausibles dans le contexte où nous nous trouvons seront, selon toute vraisemblance, situées sur l'intervalle $[0, 2]$.

4 Exemple numérique exhaustif

Dans cette section, nous présentons un exemple numérique exhaustif qui permet d'observer le fonctionnement de l'indicateur de performance décrit ci-dessus. Pour les besoins de cet exercice, un générateur de nombres aléatoires a été utilisé pour produire des données (c'est-à-dire le pourcentage de votes reçus, le montant final du budget de fonctionnement et le résultat de l'élection) pour 100 participants fictifs. Les données brutes sont présentées à la sous-section 4.1. Les valeurs numériques des trois paramètres de mise à l'échelle utilisés pour calculer respectivement $V(x)$, $B(y)$ et $E(z)$ sont quant à elles calculées à la sous-section 4.2. Le score de performance global pour chaque participant est donné à la sous-section 4.3. Enfin, à la sous-section 4.4, nous concluons en formulant quelques observations générales.

4.1 Jeu de données

Afin de pouvoir présenter un exemple d'application de l'indicateur de performance que nous avons développé, nous avons besoin d'une abondance de données à traiter. Aux fins de cet exercice, il importe peu que les données aient été obtenues dans un cadre expérimental réel ou non. Par conséquent, dans un souci de simplicité et de flexibilité, nous avons créé un ensemble de données réalistes pour 100 participants fictifs à l'aide d'un générateur de nombres aléatoires. Nous avons ensuite inséré ces données dans l'indicateur de performance afin (1) de fournir un exemple illustratif de toutes les étapes impliquées dans le calcul du score de performance

globale d'un participant donné ; et (2) de vérifier si l'indicateur de performance se comporte comme prévu.

Pour tenir compte du pourcentage de votes reçus pour nos participants fictifs à l'étude, nous avons généré 100 nombres décimaux aléatoires à partir d'une distribution normale avec une espérance de $\mu_x = 30$ et un écart type de $\sigma_x = 10$. Les résultats ont été arrondis à la première décimale. Pour la situation budgétaire, nous avons généré 100 nombres entiers aléatoires à partir d'une distribution normale avec une espérance de $\mu_y = -500$ et un écart type de $\sigma_y = 200$. Enfin, pour tenir compte du résultat des élections, nous avons utilisé un générateur aléatoire de nombres binaires pour produire 100 nombres binaires, 1 symbolisant que le participant a été réélu et 0 représentant une défaite électorale.

Le tableau ci-dessous contient les données ainsi obtenues :

#	x (%)	y	z	#	x (%)	y	z	#	x (%)	y	z
1	39.3	-685	1	35	22.1	-341	0	68	24	-380	1
2	46.4	-828	0	36	30.9	-518	1	69	20.9	-317	0
3	33.2	-565	0	37	33.5	-570	1	70	25.9	-419	1
4	31.9	-537	1	38	31	-521	1	71	26	-421	1
5	46.6	-831	1	39	46.1	-822	0	72	17.9	-257	1
6	14.3	-186	0	40	19.9	-299	1	73	23.1	-362	1
7	12.4	-148	0	41	25.4	-408	1	74	39.1	-683	0
8	23.1	-362	0	42	38.2	-664	0	75	36	-619	1
9	48.3	-867	0	43	41.8	-736	1	76	23.7	-375	0
10	22.5	-349	1	44	33.6	-572	0	77	30.4	-508	1
11	33.2	-563	1	45	25.9	-418	1	78	41.2	-724	1
12	6.6	-31	0	46	39.5	-690	1	79	34	-581	1
13	21.9	-337	0	47	32.4	-548	0	80	9	-81	1
14	47	-840	1	48	44.4	-788	1	81	5	0	1
15	40.8	-716	1	49	33.9	-579	1	82	13.8	-176	1
16	36.8	-637	0	50	40.7	-715	0	83	27.9	-457	1
17	18.4	-267	1	51	9.9	-98	0	84	23.8	-377	1
18	22.3	-347	1	52	25.3	-407	0	85	54.6	-991	1
19	22.8	-357	1	53	23.2	-364	1	86	30.2	-505	1
20	27.9	-458	0	54	43.8	-775	0	87	37.9	-657	0
21	30.7	-515	0	55	31	-521	0	88	44.8	-796	0
21	22.8	-356	0	56	27.4	-448	0	89	39.2	-684	0
23	49.4	-889	0	57	24.7	-393	0	90	25.7	-414	0
24	21.3	-326	1	58	40.9	-717	0	91	22.6	-352	0

25	43.8	-775	0		59	32.6	-553	1		92	26.7	-435	0
26	51	-921	0		60	33.7	-574	0		93	25.7	-413	0
27	39.4	-688	1		61	18.7	-273	0		94	25.1	-403	1
28	34.6	-593	0		62	31.2	-523	0		95	23.5	-370	1
29	1.7	67	1		63	35.5	-610	1		96	40.5	-371	1
30	17.2	-245	0		64	30.9	-682	0		97	20.3	-306	0
31	32.9	-558	0		65	29.3	-485	0		98	28.7	-474	0
32	25.2	-405	1		66	39.1	-682	0		99	39.3	-687	0
33	33.6	-572	0		67	38.6	-673	1		100	38.6	-687	1
34	27.8	-457	1										

TABLEAU 1: Le pourcentage de votes reçus (x), le montant final du budget de fonctionnement (y), et le résultat de l'élection (z) pour 100 participants fictifs.

4.2 Calcul des paramètres de mise à l'échelle

Tel qu'indiqué à la sous-section 3.1, les trois fonctions impliquées dans le calcul du score de performance global comportent des paramètres de mise à l'échelle dont le rôle est de garantir que les fonctions s'accordent bien entre elles, sans qu'aucune fonction n'ait un impact tellement prédominant sur le score de performance global qu'elle occulte presque complètement l'impact des deux autres fonctions.

Pour calculer ces paramètres de mise à l'échelle, il nous faut connaître la médiane ν de x et y ainsi que l'écart absolu médian par rapport à la médiane MAD_ν de x et y . La médiane est facilement obtenue en classant les valeurs de x (resp. y) dans une liste ascendante et en prenant la moyenne arithmétique des valeurs en 50^e et 51^e positions. Nous calculons ensuite, pour chaque participant, la distance (en valeur absolue) séparant son score de la médiane. Pour obtenir le MAD_ν de x (resp. y), il suffit de trier ces distances par ordre croissant puis de prendre la moyenne arithmétique des deux valeurs en 50^e et 51^e position.

Le résultat de ces calculs est présenté dans le tableau ci-dessous :

	x	y
ν	30.8 %	-511.5
MAD_ν	7.75 %	151

TABLEAU 2: La médiane et l'écart absolu médian par rapport à la médiane du pourcentage de votes reçus (x) et du montant final du budget de fonctionnement (y) pour nos 100 participants fictifs.

Pourcentage de votes reçu

Afin de calculer le paramètre de mise à l'échelle pour la fonction $B(y)$ traitant du montant final du budget de fonctionnement, nous devons d'abord effectuer un calcul pour déterminer la distance (calculée en nombre de fois l'écart absolu médian par rapport à la médiane) entre la plus petite valeur de x (à savoir 1.7%) dans notre ensemble de données et la valeur médiane de x . On peut aisément vérifier que

$$1.7 \% = \nu(x) - 3.755 \times \text{MAD}_\nu(x).$$

Nous sommes alors en mesure de trouver la valeur du paramètre de mise à l'échelle en appliquant la formule suivante :

$$k = \frac{1}{\nu(y) - 3.755 \times \text{MAD}_\nu(y)} \log \left(\frac{2}{S(\nu(x) - 3.75 \times \text{MAD}_\nu(x))} - 1 \right).$$

Une fois ce paramètre de mise à l'échelle obtenu, la fonction $B(y)$ est alors bien définie.

Résultat de l'élection

Il ne reste plus qu'à déterminer le malus m . Ce valeur s'obtient en effectuant le calcul suivant :

$$m = \frac{S(\nu(x) - \frac{1}{4} \times \text{MAD}_\nu(x)) B(\nu(y) - \frac{1}{4} \times \text{MAD}_\nu(y))}{S(\nu(x) + \frac{1}{4} \times \text{MAD}_\nu(x)) B(\nu(y) + \frac{1}{4} \times \text{MAD}_\nu(y))}.$$

Une fois ce paramètre de mise à l'échelle obtenu, la fonction $E(z)$ est alors bien définie.

Résumé

Nous présentons ci-dessous un tableau avec les valeurs numériques approchées des paramètres c , k et m :

Paramètres	Valeurs numériques approchées
c	2.564
k	0.004
m	0.695

TABLEAU 3: Valeurs numériques approchées des paramètres de mise à l'échelle c , k et m pour l'ensemble de données associé à nos 100 participants fictifs.

4.3 Calcul des indicateurs de performance globale

Le tableau suivant montre, pour chacun des 100 participants fictifs, les indicateurs de performance obtenus en introduisant leurs scores concernant le pourcentage de votes reçus, le montant final du budget de fonctionnement et le résultat de l'élection dans les fonctions $V(x)$, $B(y)$ et $E(z)$ respectivement. Enfin, les indicateurs de performance globale sont obtenus en appliquant (3.1). Ces données ont été obtenues à partir des données brutes présentées dans le tableau 1. Pour éviter la propagation inutile des erreurs d'arrondi, nous avons conservé le plus grand nombre possible de décimales tout au long des calculs et n'avons arrondi à la troisième décimale qu'à la toute fin.

#	$V(x)$	$B(y)$	$E(z)$	$F(x, y, z)$	#	$V(x)$	$B(y)$	$E(z)$	$F(x, y, z)$
1	1.065	0.164	1	0.559	51	0.255	0.829	0.695	0.528
2	1.288	0.102	0.695	0.451	52	0.663	0.384	0.695	0.562
3	0.885	0.240	0.695	0.529	53	0.606	0.434	1	0.641
4	0.848	0.262	1	0.605	54	1.205	0.122	0.695	0.468
5	1.295	0.101	1	0.508	55	0.822	0.275	0.695	0.539
6	0.369	0.683	0.695	0.560	56	0.721	0.342	0.695	0.555
7	0.320	0.745	0.695	0.549	57	0.647	0.400	0.695	0.565
8	0.603	0.436	0.695	0.568	58	1.114	0.148	0.695	0.485
9	1.351	0.090	0.695	0.439	59	0.868	0.249	1	0.600
10	0.587	0.452	1	0.643	60	0.899	0.233	0.695	0.526
11	0.885	0.241	1	0.598	61	0.485	0.553	0.695	0.571
12	0.169	0.945	0.695	0.481	62	0.828	0.273	0.695	0.540
13	0.571	0.467	0.695	0.570	63	0.952	0.208	1	0.583
14	1.308	0.098	1	0.505	64	0.819	0.277	1	0.610
15	1.111	0.148	1	0.548	65	0.774	0.306	0.695	0.548
16	0.990	0.191	0.695	0.509	66	1.059	0.166	0.695	0.496
17	0.477	0.561	1	0.645	67	1.044	0.170	1	0.562

18	0.582	0.454	1	0.642		68	0.628	0.415	1	0.639
19	0.595	0.442	1	0.641		69	0.544	0.493	0.695	0.571
20	0.735	0.332	0.695	0.553		70	0.680	0.371	1	0.632
21	0.813	0.280	0.695	0.541		71	0.682	0.369	1	0.632
22	0.595	0.443	0.695	0.568		72	0.464	0.575	1	0.644
23	1.388	0.083	0.695	0.432		73	0.603	0.436	1	0.641
24	0.555	0.481	1	0.644		74	1.059	0.165	0.695	0.495
25	1.205	0.122	0.695	0.468		75	0.966	0.203	1	0.581
26	1.443	0.075	0.695	0.422		76	0.619	0.421	0.695	0.566
27	1.068	0.162	1	0.558		77	0.805	0.286	1	0.613
28	0.925	0.220	0.695	0.521		78	1.123	0.144	1	0.545
29	0.044	1.118	1	0.365		79	0.908	0.228	1	0.592
30	0.445	0.593	1	0.642		80	0.231	0.858	1	0.583
31	0.876	0.245	0.695	0.531		81	0.128	1.000	1	0.504
32	0.660	0.387	1	0.634		82	0.356	0.699	1	0.629
33	0.896	0.235	0.695	0.527		83	0.735	0.333	1	0.625
34	0.732	0.333	1	0.624		84	0.622	0.418	1	0.639
35	0.576	0.462	0.695	0.570		85	1.571	0.059	1	0.452
36	0.819	0.277	1	0.610		86	0.799	0.288	1	0.613
37	0.893	0.236	1	0.595		87	1.023	0.179	0.695	0.503
38	0.822	0.275	1	0.609		88	1.236	0.114	0.695	0.461
39	1.278	0.104	0.695	0.453		89	1.062	0.164	0.695	0.495
40	0.517	0.517	1	0.644		90	0.674	0.377	0.695	0.561
41	0.666	0.383	1	0.634		91	0.590	0.448	0.695	0.569
42	1.032	0.175	0.695	0.501		92	0.702	0.355	0.695	0.557
43	1.142	0.139	1	0.541		93	0.674	0.378	0.695	0.562
44	0.896	0.235	0.695	0.527		94	0.658	0.389	1	0.635
45	0.680	0.373	1	0.633		95	0.614	0.427	1	0.640
46	1.071	0.161	1	0.557		96	1.102	0.425	1	0.777
47	0.862	0.253	0.695	0.533		97	0.528	0.507	0.695	0.571
48	1.224	0.117	1	0.523		98	0.757	0.316	0.695	0.550
49	0.905	0.230	1	0.592		99	1.065	0.163	0.695	0.494
50	1.108	0.149	0.695	0.486		100	1.044	0.171	1	0.563

TABLEAU 4: Indicateurs de performance partiels et global pour nos 100 participants fictifs.

4.4 Analyse et discussion

Il convient maintenant de procéder à l'analyse des données présentées dans le Tableau 4.

Le tableau suivant présente les mesures de tendance centrale et de dispersion statistique des indicateurs de performance partiels (à savoir $V(x)$ et $B(y)$) et global (c'est-à-dire $F(x, y, z)$) pour nos 100 participants fictifs.

	$V(x)$	$B(y)$	$F(x, y, z)$
μ	0.812	0.336	0.562
σ	0.301	0.207	0.065
Q_0 (min)	0.044	0.059	0.365
Q_1	0.605	0.171	0.526
Q_2 (ν)	0.816	0.283	0.562
Q_3	1.048	0.434	0.611
Q_4 (max)	1.571	1.118	0.777
MAD_ν	0.214	0.120	0.048
éI	0.442	0.264	0.085

TABLEAU 5: Mesures de tendance centrale et de dispersion statistique pour les indicateurs de performance partiels et global de nos 100 participants fictifs.

Voici un graphique de la distribution des indicateurs de performance globale.

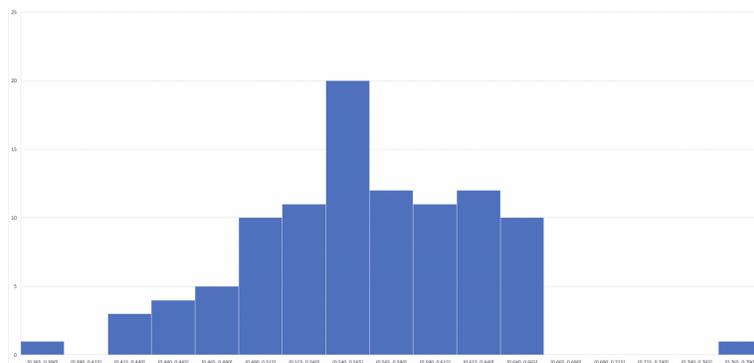


FIGURE 5 – Distribution des indicateurs de performance pour nos 100 participants fictifs.

L'un des aspects les plus visuellement frappants de la Figure 5 est la queue de gauche plus longue, suggérant une asymétrie négative. Cette impression est confirmée par le calcul de l'asymétrie non paramétrique et du coefficient de moment de Fisher de l'asymétrie, car ces deux mesures de l'asymétrie s'avèrent être des nombres négatifs de faible module. En effet, en

utilisant la formule de l'asymétrie non paramétrique, définie comme $(\mu - \nu)/\sigma$, on obtient que l'asymétrie est de $-0,011$. Quant au coefficient d'asymétrie du moment de Fisher, défini comme

$$E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right],$$

il est de $-0,136$. Cette légère asymétrie à gauche n'est pas surprenante étant donné la nature dichotomique de la fonction qui traite du résultat de l'élection. En effet, le caractère punitif de la fonction $E(z)$ entraîne une réduction de la performance globale de tous les participants qui ont subi une défaite électorale (cela se traduit graphiquement par un déplacement vers la gauche d'une large part des participants à l'étude). D'ailleurs, la plupart des participants ayant subi une défaite électorale se situent dans les premier et deuxième quartiles. Le malus n'est cependant pas d'une ampleur telle qu'il donne lieu à deux distributions disjointes. En effet, d'une part, certains participants battus (par exemple les participants # 61, 69, 98) ont réalisé des performances supérieures à la moyenne et à la médiane. D'autre part, certains participants élus (par exemple, les participants # 29, 81 et 85) ont obtenu des scores globaux qui les placent parmi les moins performants. Par conséquent, comme souhaité, le résultat de l'élection ne prédomine pas au point de masquer entièrement l'impact des autres variables. De le même ordre d'idée, on peut facilement vérifier qu'une surperformance selon n'importe lequel des indicateurs n'est jamais suffisante à elle seule pour qu'un participant se démarque (il suffit d'observer des participants comme #12, 14, 26, 81, 85 pour s'en rendre compte). Mais une performance nettement supérieure (resp. inférieure) à la médiane selon tous les indicateurs se traduit, comme prévu, par une performance globale nettement supérieure (resp. inférieure) à la médiane ; la forme multiplicative de l'indicateur de performance globale tend alors à amplifier considérablement l'écart avec la médiane (voir le participant # 96).

5 Discussion

Les environnements de décision du monde réel sont complexes. Il est particulièrement ardu d'examiner les processus cognitifs de prise de décision en utilisant l'expérimentation directe par essai-erreur, en raison des graves conséquences potentielles qu'un tel exercice peut avoir dans la vie réelle. Les micromondes simulés par ordinateur offrent une opportunité unique d'étudier la

prise de décision dans des contextes qui capturent les propriétés de la complexité. Néanmoins, il demeure essentiel de jumeler l'utilisation d'un micromonde à un indicateur capturant la complexité dynamique et la difficulté qu'encourt le décideur pour gérer cette même complexité, notamment lorsqu'il s'efforce d'harmoniser des objectifs contradictoires.

Cette étude avait pour objectif d'évaluer le potentiel d'une mesure ouverte de la performance à une tâche de RPC à travers l'utilisation du jeu de simulation D3 comme micromonde de recherche capturant les attributs propres aux environnements complexes du monde réel. La mesure de la performance utilisée devait tenir compte de la complexité dynamique en considérant la *polytely*. La qualité psychométrique de l'indicateur développé a été évaluée sur la base d'une simulation complète de cas générés aléatoirement, permettant ainsi de juger de l'applicabilité du micromonde comme méthode d'évaluation de la performance à une tâche de RPC. Les résultats démontrent d'abord que D3 offre des conditions psychométriques satisfaisantes pour l'évaluation de la RPC. En effet, la simulation capture les différentes propriétés de la complexité tout en permettant l'extraction de données comportementales à partir d'indicateurs présents dans le jeu. Ces indicateurs offrent ensuite la possibilité de tenir compte dans la modélisation de la performance de la difficulté pour le joueur de concilier des objectifs entrant en contradiction les uns avec les autres. En outre, l'étude contribue à la littérature sur la RPC en suggérant qu'il est utile de tenir compte de la *polytely* dans l'évaluation de la performance à une tâche de décision dynamique.

Implications théoriques

La littérature reconnaît un certain nombre de propriétés aux environnements complexes. Les chercheurs s'entendent généralement pour dire que les environnements de décision du monde réel sont complexes, connectés, dynamiques, opaques (partiellement ou entièrement), et *polytéliques* (p. ex., [20, 37, 40, 47]). La présente étude a proposé une méthode de capture de la complexité dynamique dans la mesure de la performance à une tâche de décision dans un environnement complexe. L'évaluation de la dépendait de la manière dont l'environnement de tâche était capable de simuler les propriétés de la complexité, et de la capacité de la procédure mathématique à produire un indicateur de performance reflétant la difficulté de concilier des

objectifs contradictoires, tout en permettant la comparaison des scores entre les sujets, sur la base d'un seul et même indicateur numérique. Spécifiquement, l'indicateur a permis d'attribuer de manière cohérente une valeur numérique à chaque triplet composé du pourcentage de votes obtenus, de la situation budgétaire, et du résultat de l'élection. La cohérence de l'indicateur de performance a été reconnue puisque ce dernier était en mesure de reproduire le comportement attendu lorsque deux des trois paramètres étaient fixes et que le troisième variait librement sur l'ensemble du domaine. Cette cohérence a également été observée lorsque les trois paramètres tendaient simultanément vers des valeurs extrêmes ou centrales. La qualité psychométrique de l'indicateur de performance suggère que la procédure employée offre une base robuste pour comparer les performances des participants dont le triplet de sortie présente un profil plus hétérogène.

Implications pratiques

L'opérationnalisation de la complexité à travers des micromondes tels que D3 permet d'évaluer le comportement décisionnel face à des problèmes complexes. Un indicateur de performance capable de tenir compte de la difficulté pour les individus de concilier des objectifs contradictoires dans le contrôle de systèmes complexes constitue un pas de plus vers une évaluation valide de la performance dans une tâche de RPC. Cette étude est un plaidoyer en faveur de la prise en compte des propriétés de la complexité, non seulement dans le choix du micromonde de recherche utilisé pour étudier la RPC, mais aussi dans l'évaluation proprement dite de la performance, de sorte que la mesure soit le résultat direct du traitement de l'information complexe et dynamique. Au-delà de la présente étude, ce type de mesure multidimensionnelle de la performance pourrait être transféré à un large éventail de situations dans une variété de domaines, notamment l'évaluation neuropsychologique clinique [41], la sélection du personnel [36], et l'éducation [17].

Une formule pluridimensionnelle

1. Les fonctions qui modélisent correctement le changement souhaité lorsque l'on autorise des variations comportementales (*ceteris paribus*) du pourcentage de votes reçus ou de

la situation budgétaire ne sont en aucun cas uniques. Par exemple, plutôt que d'utiliser une tangente hyperbolique inverse mise à l'échelle, nous aurions pu utiliser une fonction tangente mise à l'échelle ou une exponentielle décalée et mise à l'échelle de $\frac{1}{1-x}$. Quant à la fonction logistique, elle pourrait être remplacée par une autre fonction sigmoïde telle qu'une tangente hyperbolique décalée et mise à l'échelle, une fonction d'erreur de Gauss ou une fonction de Gompertz. Un tel changement modifierait la valeur numérique associée à l'indicateur de performance global d'un participant, mais les tendances observables resteraient en grande partie les mêmes.

2. Le fait que le calcul des paramètres de mise à l'échelle c et k ainsi que le calcul du malus impliquent des informations provenant de la distribution des données s'avère être une des forces de notre modèle puisqu'il permet un maillage harmonieux des fonctions $V(x)$, $B(y)$ et $E(z)$. En effet, en procédant de la sorte, aucune de ces trois fonctions n'a un impact tellement prédominant sur la performance globale que l'effet des autres paramètres est masqué. Cette approche a cependant l'inconvénient de rendre relatif l'indicateur de performance globale d'un participant donné. En effet, la valeur associée à une performance n'a de sens que dans le cadre d'une comparaison avec les performances d'autres participants appartenant au même échantillon. Notons que cela n'empêche pas de faire des comparaisons entre les performances de participants appartenant à des échantillons différents. Pour ce faire, il suffit de fusionner les deux sous-échantillons et de calculer une fois pour toutes les valeurs numériques des paramètres c , k et m .
3. Le principal avantage de l'utilisation de la médiane plutôt que de la moyenne (et de la variation absolue de la médiane plutôt que de l'écart-type) est que la première est plus résistante aux valeurs aberrantes que la seconde. Toutefois, si, pour une distribution de données donnée, la variance relative de la médiane est beaucoup plus grande que la variance de la moyenne, et si cela devait devenir un problème, on pourrait alors remplacer la médiane par une moyenne tronquée, ce qui implique le calcul de la moyenne après avoir écarté une quantité égale de données aux extrémités haute et basse de la distribution. Une moyenne tronquée utilise plus d'informations de la distribution ou de l'échantillon que la médiane (ce qui la rend moins sensible au bruit de la médiane), tout en restant moins affectée par les valeurs aberrantes que la moyenne. Dans ce cas, une

mesure appropriée de la dispersion serait l'écart absolu de la médiane par rapport à la moyenne tronquée.

4. Afin de tester notre hypothèse selon laquelle l'asymétrie à gauche de la distribution des indicateurs de performance est inhérente au modèle, nous avons généré deux grands ensembles de données composés de 10 000 participants fictifs à l'étude. Pour le premier, nous avons utilisé les mêmes moyennes et écarts types que précédemment, c'est-à-dire $\mu_x = 30$, $\sigma_x = 10$, $\mu_y = -500$, et $\sigma_y = 200$. Pour des raisons évidentes, les quelques valeurs négatives du pourcentage de votes reçus ont été corrigées (c'est-à-dire remplacées par de nouvelles données générées aléatoirement). Une fois de plus, nous avons obtenu une distribution avec une asymétrie non paramétrique négative (-0.05) et un coefficient d'asymétrie du moment de Fisher modérément négatif (-0.52). Pour la seconde, nous avons utilisé une distribution entièrement équilibrée, c'est-à-dire $\mu_x = 50$, $\sigma_x = 12$, $\mu_y = 0$, et $\sigma_y = 250$. Il est intéressant de noter que nous avons obtenu une fois de plus une distribution avec une asymétrie non paramétrique négative (-0.03), et un coefficient d'asymétrie du moment de Fisher légèrement négatif (-0.12).

Limites et poursuite de la recherche

Il est inévitable que toute fonction qui attribue, dans un type de contexte comme le nôtre, une valeur numérique unique à un n -tuple de données expérimentales comporte un certain degré de subjectivité, car une telle fonction est, par nature, un dispositif mathématique permettant de pondérer la contribution de chacune des composantes, et il existe plusieurs façons différentes d'effectuer une telle pondération. Bien que les procédures de mise à l'échelle que nous avons développées pour notre indicateur de performance utilisent des informations tirées de l'ensemble de l'échantillon afin de réduire l'arbitraire, un certain degré de subjectivité subsiste dans le choix des ajustements d'échelles (ceci est particulièrement visible dans le cas du calcul du malus). Ce degré de subjectivité a inévitablement un impact sur l'ordre de certains participants.

Il convient également de noter que notre choix d'utiliser la médiane et l'écart absolu médian par rapport à la médiane plutôt que la moyenne et l'écart type dans le calcul des paramètres de mise à l'échelle présente l'avantage important de rendre le calcul insensible aux valeurs aberrantes.

D'autre part, la médiane et l'écart absolu de la médiane sont considérablement affectés par (1) l'ajout ultérieur de nouvelles données ; (2) les fluctuations aléatoires des données. L'utilisation de cet indicateur de performance dans un contexte expérimental révélera si cet inconvénient s'avère problématique. Dans ce cas, il peut être nécessaire de modifier la manière de calculer les paramètres de mise à l'échelle (comme, par exemple, en faisant un compromis entre la moyenne et la médiane en considérant une moyenne tronquée).

Références

- [1] Elena BAKHANOVA, Jaime A GARCIA, William L RAFFE et Alexey VOINOV : Targeting social learning and engagement : What serious games and gamification can offer to participatory modeling. *Environmental Modelling & Software*, 134:104846, 2020.
- [2] Stephan BARTHOLDY et Ulrike KIPMAN : Influences of reasoning and achievement motivation on complex problem solving in a new microworld operationalization. *Journal of Global Education and Research*, 3(2):141–157, 2019.
- [3] Benoît BÉCHARD : Tout simplement humain : une étude de la complexité politique. 2023.
- [4] Benoît BÉCHARD, Helen HODGETTS, Frédéric MORNEAU-GUÉRIN, Mathieu OUMET et Sébastien TREMBLAY : Political complexity and the pervading role of ideology in policy-making. *Journal of Dynamic Decision Making*, 2023.
- [5] Walter R. BOOT : Video games as tools to achieve insight into cognitive processes., 2015.
- [6] Berndt BREHMER : Man as a stabiliser of systems : From static snapshots of judgement processes to dynamic decision making. *Thinking & Reasoning*, 2(2-3):225–238, 1996.
- [7] Ernst DIEHL et John D. STERMAN : Effects of feedback complexity on dynamic decision making. *Organizational behavior and human decision processes*, 62(2):198–215, 1995.
- [8] Emanuel DONCHIN : Video games as research tools : The Space Fortress game. *Behavior Research Methods, Instruments, & Computers*, 27(2):217–223, 1995.
- [9] Dietrich DÖRNER : *Problemlösen*. Bundesamt für Wehrtechnik und Beschaffung, 1975.
- [10] Dietrich DÖRNER : The logic of failure : Recognizing and avoiding error in complex situations. *Reading, MA, USA*, 1997.

- [11] Dietrich DÖRNER et Joachim FUNKE : Complex problem solving : What it is and what it is not. *Frontiers in psychology*, 8:1153, 2017.
- [12] Dietrich DÖRNER et C Dominik GÜSS : Human error in complex problem solving and dynamic decision making : A taxonomy of 24 errors and a theory. *Computers in Human Behavior Reports*, 7:100222, 2022.
- [13] Dietrich DÖRNER, Heinz W KREUZIG, Franz REITHER et Thea STÄUDEL : Lohhausen : Vom Umgang mit Unbestimmtheit und Komplexität. 1983.
- [14] Dietrich DÖRNER et Franz REITHER : Über das Problemlösen in sehr komplexen Realitätsbereichen. *Zeitschrift für experimentelle und angewandte Psychologie : Organ der Deutschen Gesellschaft für Psychologie*, 25(4):527–551, 1978.
- [15] Michael ENGELHART, Joachim FUNKE et Sebastian SAGER : A decomposition approach for a new test-scenario in complex problem solving. *Journal of Computational Science*, 4(4):245–254, 2013.
- [16] Joachim FUNKE : Problem solving. *The psychology of human thought : An introduction*, pages 155–176, 2019.
- [17] Joachim FUNKE : It requires more than intelligence to solve consequential world problems. *Journal of Intelligence*, 9(3):38, 2021.
- [18] Joachim FUNKE : Be prepared for the complexities of the Twenty-First Century! *In The Palgrave Handbook of Transformational Giftedness for Education*, pages 171–180. Springer, 2022.
- [19] Joachim FUNKE, Andreas FISCHER et Daniel V HOLT : When less is less : Solving multiple simple problems is not complex problem solving,Äÿa comment on greiff et al.(2015). *Journal of Intelligence*, 5(1):5, 2017.
- [20] Joachim FUNKE et Peter A. FRENCH : Complex problem solving : The European perspective—10 years after. *In Learning to solve complex scientific problems*, pages 25–48. Routledge, 2017.
- [21] Juan Martín GARCÍA : *Theory and practical exercises of system dynamics : Modeling and simulation with Vensim PLE*. Juan Martin Garcia, 2020.
- [22] Cleotilde GONZALEZ, Polina VANYUKOV et Michael K. MARTIN : The use of microworlds to study dynamic decision making. *Computers in human behavior*, 21(2):273–286, 2005.

- [23] Wayne D. GRAY : Simulated task environments : The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly*, 2(2):205–227, 2002.
- [24] Samuel GREIFF, Matthias STADLER, Philipp SONNLEITNER, Christian WOLFF et Romain MARTIN : Sometimes less is more : Comparing the validity of complex problem solving measures. *Intelligence*, 50:100–113, 2015.
- [25] Todd M. GURECKIS et Bradley C. LOVE : Short-term gains, long-term pains : How cues about state aid learning in dynamic environments. *Cognition*, 113(3):293–313, 2009.
- [26] Graeme S. HALFORD, William H. WILSON et Steven PHILLIPS : Processing capacity defined by relational complexity : Implications for comparative, developmental, and cognitive psychology. *Behavioral and brain sciences*, 21(6):803–831, 1998.
- [27] Cliff HARRIS : Democracy 3 : Institute of effective government., 2013.
- [28] Christoph Nils HERDE, Sascha WÜSTENBERG et Samuel GREIFF : Assessment of complex problem solving : What we know and what we don't know. *Applied Measurement in Education*, 29(4):265–277, 2016.
- [29] Peter J. HUBER : *Robust statistics*. John Wiley, 1981.
- [30] David KIRSH : A few thoughts on cognitive overload. 2000.
- [31] Daniel LAFOND, Michel B. DUCHARME, Jean-François GAGNON et Sébastien TREMBLAY : Support requirements for cognitive readiness in complex operations. *Journal of Cognitive Engineering and Decision Making*, 6(4):393–426, 2012.
- [32] Daniel LAFOND, Jean-François GAGNON, Marie-Ève ST-LOUIS, Sylvain PRONOVOST, Michel B DUCHARME et Sébastien TREMBLAY : Decision heuristics and human performance in a policy management simulation. *In 32nd International Conference of the System Dynamics Society : Good Governance in a Complex World*, pages 1–13, 2014.
- [33] Paul A. LANGLEY et John D. W. MORECROFT : Performance and learning in a simulation of oil industry dynamics. *European Journal of Operational Research*, 155(3):715–732, 2004.
- [34] Christophe LEYS, Christophe LEY, Olivier KLEIN, Philippe BERNARD et Laurent LICATA : Detecting outliers : Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology*, 49(4):764–766, 2013.

- [35] Christian Sebastian LOH et Yanyan SHENG : Measuring expert performance for serious games analytics : From data to insights. *In Serious Games Analytics*, pages 101–134. Springer, 2015.
- [36] Jakob MAINERT, Christoph NIEPEL, Kevin R MURPHY et Samuel GREIFF : The incremental contribution of complex problem-solving skills to the prediction of job level, job complexity, and salary. *Journal of Business and Psychology*, 34:825–845, 2019.
- [37] Donella MEADOWS, Jorgen RANDERS et Dennis MEADOWS : *Limits to growth : The 30-year update*. Chelsea Green Publishing, 2004.
- [38] Alan MILLER : Tunnel vision in environmental management. *The Environmentalist*, 2(3):223–231, 1982.
- [39] Ioannis NIKOLAOU, Konstantina GEORGIU et Vasiliki KOTSASARLIDOU : Exploring the relationship of a gamified assessment with performance. *The Spanish Journal of Psychology*, 22:E6, 2019.
- [40] Magda OSMAN : Controlling uncertainty : A review of human behavior in complex dynamic environments. *Psychological bulletin*, 136(1):65, 2010.
- [41] Sonia PACKWOOD, Helen M HODGETTS et Sébastien TREMBLAY : A multiperspective approach to the conceptualization of executive functions. *Journal of clinical and experimental neuropsychology*, 33(4):456–470, 2011.
- [42] Mark PAICH et John D. STERMAN : Boom, bust, and failures to learn in experimental markets. *Management science*, 39(12):1439–1458, 1993.
- [43] Cliff Harris POSITECH GAMES : Democracy : Take control of your country., [2005, 2007, 2013] 2021.
- [44] Genaro REBOLLEDO-MENDEZ, Katerina AVRAMIDES, Sara DE FREITAS et Kam MEMARZIA : Societal impact of a serious game on raising public awareness : the case of FloodSim. *In Proceedings of the 2009 ACM SIGGRAPH symposium on video games*, pages 15–22, 2009.
- [45] Georgios RIGAS, Eva CARLING et Berndt BREHMER : Reliability and validity of performance measures in microworlds. *Intelligence*, 30(5):463–480, 2002.
- [46] Margarida ROMERO, Mireia USART et Michela OTT : Can serious games contribute to developing and sustaining 21st century skills? *Games and culture*, 10(2):148–177, 2015.

- [47] John D STERMAN : System dynamics modeling : tools for learning in a complex world. *California management review*, 43(4):8-25, 2001.
- [48] John D. STERMAN : Learning from evidence in a complex world. *American journal of public health*, 96(3):505-514, 2006.
- [49] Benjamin G. STOKES : Videogames have changed : Time to consider serious games? *Development Education Journal*, 11(3):12, 2005.
- [50] Gabrielle TEYSSIER-ROBERGE, Helen HODGETTS, Benoît BÉCHARD, Francois-Bernard MALO et Sébastien TREMBLAY : A microworld simulation for the assessment of complex problem-solving skills in personnel selection. Manuscript under review in *Frontiers in Psychology*, 2023.
- [51] Frederic VESTER : Ecopolity : A cybernetic environmental simulation game. *English/German version*, 2010.