

L'architecture du feedback

Pourquoi des distributions d'erreurs égales produisent des dérives institutionnelles inégales

Résumé. Les théories de l'apprentissage organisationnel demandent comment une institution apprend, imparfaitement, du feedback qu'elle reçoit. Les théories de l'oubli organisationnel demandent comment elle perd un savoir qu'elle a un jour détenu. Cet article se situe en amont des deux. Il demande quelles erreurs laissent une trace, et soutient que ce filtre est fixé par l'architecture du feedback, avant tout apprentissage et avant tout oubli. La thèse centrale est conditionnelle, et la condition fait sa force : à distributions d'erreurs initiales égales, les politiques institutionnelles divergent lorsque leurs architectures de feedback imposent des forces contraignantes différentes à la boucle de décision locale. Les politiques n'évoluent pas à partir des seules distributions d'erreurs. Elles évoluent à travers l'architecture qui transforme certaines erreurs en feedback contraignant et en laisse d'autres muettes. La porte de fiabilité est le terrain sur lequel la thèse est démontrée. Le cliquet d'observabilité, traité dans une ligne antérieure de ce travail comme le mécanisme central, en est ici un corollaire.

1. Un principe, et un changement de discipline qu'il ne cache pas

La porte de fiabilité a été proposée, dans la livraison précédente¹, comme un dispositif constructif : un lieu où le compromis entre sûreté et utilité pouvait être encodé par l'architecture plutôt que laissé à la discrétion. Cette proposition tient. Le présent article lui fait subir l'opération inverse. Il retourne le dispositif contre lui-même et le regarde fonctionner, car une porte qui fonctionne parfaitement peut néanmoins produire une pathologie, et cette pathologie est le sujet.

La thèse d'abord. ***Une politique de fiabilité ne dérive pas selon la fréquence de ses erreurs ni selon leur coût, pris comme distribution. Elle dérive selon la façon dont son architecture distribue le feedback*** : quelles erreurs renvoient un signal assez contraignant, et assez tôt, pour peser sur la décision, et lesquelles ne renvoient rien à l'instant qui compte. Fixez la distribution des erreurs à l'identique sur deux portes, et elles

¹ J. Vétillard, « Le modèle n'a jamais été l'objet », Twingital Institute, 12 juillet 2026. Le gate y est posé dans sa fonction constructive : dispositif inline rendant, au point d'irréversibilité, l'un de trois verdicts (blocage, réparation, autorisation) avant l'engagement, plutôt qu'audit ex post laissé à la discrétion. Le présent article retourne ce dispositif contre lui-même.

pourront néanmoins dériver différemment, car la politique se meut par la forme du feedback que ses erreurs produisent, non par leur nombre.

Le compte rendu classique n'y est pas aveugle, et prétendre le contraire serait faux. Un modèle de coût peut porter des coûts de détection, des délais, des probabilités de découverte, des coûts différés, des asymétries d'information. Ce qu'il ne fait généralement pas, c'est traiter l'asymétrie du feedback comme un mécanisme architectural autonome de dérive. Il classe l'asymétrie comme un paramètre exogène, un nombre remis de l'extérieur, plutôt que comme une propriété structurelle de la porte qui engendre la dérive de l'intérieur. Les disciplines oublient rarement des variables ; elles les rangent en annexe. La contribution consiste à déplacer celle-ci de l'annexe vers le mécanisme, et à la lier à l'architecture qui la produit.

Cet article change de discipline chemin faisant, de la gouvernance de la fiabilité de l'IA vers une proposition sur la manière dont les institutions déplacent leurs politiques sous feedback asymétrique, et le changement est délibéré. Trois assertions sont tenues séparées par leur statut.

- Le *principe*, selon lequel à distribution d'erreurs égale l'architecture du feedback gouverne la divergence des politiques, est général et conjectural,
- La *loi*, sa spécialisation falsifiable aux systèmes gouvernés par une boucle de décision locale, est démontrée sur le terrain borné des portes de fiabilité de l'IA,
- L'*extension* à d'autres domaines est nommée comme conjecture.

La théorie est, à ce stade, plus mûre que sa preuve, laquelle se réduit à une seule instance de fiabilité, à un corpus de données cliniques recueillies à d'autres fins, et à un ensemble d'adjacences dans la littérature. C'est une condition courante, énoncée une fois. L'expérience qui ferait passer la loi de l'état énoncé à l'état testé est spécifiée en Section 6, et n'est pas réalisée ici.

2. Trois primitives, et une frontière

Une théorie est stable quand elle repose sur peu d'objets et dérive le reste. La discipline consiste à réduire, et à fixer, avant de réduire, l'unique frontière sur laquelle l'argument pivote. ***Dans cet article, l'architecture du feedback désigne le mécanisme sociotechnique local qui transforme une décision et ses conséquences en un signal adressé à la boucle de politique.*** Un audit ou une régulation est à l'intérieur de l'architecture lorsqu'il fait partie de ce mécanisme local, et à l'extérieur lorsqu'il n'en fait pas partie. La frontière peut inclure ou exclure un dispositif donné selon le système étudié, mais elle est fixée pour la durée d'une analyse, afin qu'endogène et exogène ne deviennent pas déplaçables au gré des besoins.

L'argument se reconstruit sur trois primitives.

- La première est la politique, Π , l'ensemble des paramètres institutionnels qui gouvernent les décisions de la porte, l'objet qui dérive,
- La deuxième est la force contraignante d'un signal de feedback, B, la contrainte qu'un signal exerce sur la boucle de décision locale,
- La troisième est la mémoire institutionnelle, M, la référence formelle qui fait foi.

De ces trois, le reste se dérive et ne doit jamais être traité comme fondamental.

- Le *Requis*, la charge de preuve qu'une politique exige à un port, est une projection de Π ,
- Le coût relatif de transformation est une fonction de Π , M et B,
- *Bruyant* et *Silencieux* sont deux régimes de B,
- Le *déclassement* est une transformation de Π ,
- Le *cliquet* est une propriété dynamique de Π sous une certaine structure de M. Aucune quatrième primitive n'est admise.

B demande du soin, car c'est l'objet le plus exposé au reproche d'être une notation commode. Il ne peut pas se construire à partir d'une liste de parties incommensurables, la latence dans le temps, l'attribution dans l'information, la récurrence dans les comptes, le coût en monnaie, sans règle de combinaison ; ces quatre-là sont ses déterminants, non ses coordonnées, et B est monotone dans les directions évidentes. Il ne peut pas non plus être une propriété du signal, de l'architecture et de l'organisation à la fois, énoncée de façon interchangeable, et la résolution est une distinction interne à B plutôt qu'une quatrième primitive.

Appelons B_A la *capacité contraignante architecturale*, la contrainte que l'architecture seule pourrait imposer, présente partout où l'architecture est présente et orientée dans un seul sens.

Appelons B_R la *force contraignante réalisée*, ce que cette capacité devient dans une organisation donnée à un moment donné, relevée ou abaissée par un droit de contournement ou une autorité laxiste. B_A est endogène à l'architecture ; B_R est relationnelle, $B_R = B(A, S, O, t)$. L'architecture possède la capacité, non la réalisation, et l'article n'attribue à l'architecture que la capacité.

Définir B par son effet exige une prise qui ne soit pas l'effet qu'il prédit. La prise candidate est le coût de suppression, et il doit être normalisé sous peine de mesurer autre chose. Le coût de suppression brut confond la force du signal avec la rigidité de la gouvernance qui l'entoure : une alerte faiblement contraignante peut coûter cher à faire taire dans une maison bureaucratique, et une alerte fortement contraignante peut être réduite au silence en deux clics dans une maison laxiste. La quantité utilisable est le coût de suppression divisé par le coût générique de modification de la politique dans cette organisation, $C_{s^*} = C_s / C_{\text{change}}$, ce qui retire l'enveloppe administrative et laisse la part attribuable au signal. Même normalisée, elle demeure un proxy de la contrainte

réalisée, non une lecture directe de B_A, et elle est employée comme tel. L'expérience de la Section 6 est ce qui donnerait à B un contenu empirique en dépit de l'impureté.

3. De la force contraignante au moteur de coût, par médiation

Pourquoi un signal contraignant produit-il une dégradation du *Requis* plutôt qu'une amélioration ? La flèche est réelle mais non directe. Elle est médiatisée, et c'est dans la médiation que l'argument gagne sa crédibilité.

Une erreur bruyante, un faux BLOCK ou un REPAIR qui arrête le clinicien en temps réel, produit un signal contraignant sur la boucle. Le signal produit une pression de correction. La pression élève le coût perçu du maintien de la ligne, du maintien du *Requis* là où il se tient tandis que la friction persiste. Ce coût perçu entre dans un choix, et le choix n'est pas binaire. Entre relever la barre et l'abaisser s'étend une famille de réponses de confinement : suspendre le système, restreindre son domaine, escalader le cas vers un humain, revenir au traitement manuel, réduire le volume envoyé à travers la porte. Le confinement maintient le *Requis* en place tout en limitant l'exposition. Le cliquet n'apparaît pas simplement lorsque la dégradation coûte moins cher que la promotion, mais lorsque la dégradation coûte moins cher que la promotion et moins cher que toute réponse de confinement disponible. Là où un confinement bon marché existe, la friction bruyante est soulagée sans toucher au *Requis*. Là où il n'existe pas, la dégradation est le chemin de moindre coût, et la dérive est le gradient de coût, suivi.

C'est le point à retenir. L'asymétrie n'est pas une défaillance du personnel. C'est la forme du champ d'incitations que l'architecture construit. ***Un acteur parfaitement rationnel dégrade le Requis sous ce champ, car réduire au silence l'erreur bruyante est bon marché et récompensé aussitôt par la chute de la friction, tandis que relever la barre contre l'erreur silencieuse est coûteux et récompensé par rien que la boucle puisse sentir.***

4. La mémoire par non-inscription

Voici le centre de l'argument, et le point où il se sépare le plus nettement de ce qui le précède.

Une décision locale, prise une fois, n'est pas encore une norme. Un seuil abaissé un mardi est une exception, aussi longtemps qu'il reste une exception au registre. Il devient norme à un moment précis : lorsque l'état dégradé est inscrit comme la référence qui fait foi, lorsque l'attestation de classe trois remplace la vérification de classe un comme l'artefact au regard duquel la conformité future est jugée. La mémoire institutionnelle

opère cette inscription, et elle est donc le niveau où un choix local se durcit en règle permanente.

L'assertion la plus forte porte sur ce que la mémoire n'inscrit pas. Une institution n'enregistre pas ce qui lui arrive. Elle enregistre ce qui laisse une trace. L'erreur bruyante en laisse une par construction : elle a bloqué quelqu'un, généré un contournement, produit un artefact qui entre au registre. L'erreur silencieuse ne laisse rien. Un faux ALLOW, passé à l'exécution contre un angle mort, ne produit ni interruption ni document, et n'entre pas en mémoire parce qu'il n'a jamais rien produit que la mémoire puisse retenir. Le biais de la mémoire institutionnelle n'est donc ni décroissance ni amnésie. C'est une non-inscription. ***Le registre penche vers le bruyant non parce que le silencieux a été oublié mais parce que le silencieux n'a jamais été écrit.*** La proposition qui en résulte est, de tout ce qui figure ici, la plus susceptible de survivre à l'article : une organisation ne peut pas apprendre institutionnellement d'un événement qui n'entre jamais dans son système de preuve, non parce qu'elle l'a choisi, ni parce qu'elle a oublié, mais parce qu'il n'y avait rien à apprendre là où l'apprentissage est conservé.

Confrontez cela à deux corpus voisins, car la ressemblance est assez proche pour être prise pour une identité. L'apprentissage organisationnel, d'après Levitt et March (1988), prend le signal reçu comme donné et étudie sa mauvaise interprétation. L'oubli organisationnel, d'après de Holan et Phillips (2004), étudie la perte d'un savoir une fois acquis. La non-inscription est en amont des deux : ni une mauvaise lecture d'un signal disponible, ni la perte d'un signal une fois détenu, mais l'absence structurelle du signal au seuil du registre, une absence que l'architecture décide à l'avance par la force contraignante qu'elle assigne. Une précision que le concept exige : une trace peut être documentaire, comportementale, métrique, juridique, narrative, ou implicite dans une routine, et une théorie générale de la mémoire les admettrait toutes. Cet article n'a pas besoin de toutes. Sur son terrain borné, le système de fiabilité régulé, M est la référence formelle qui fait foi, l'artefact d'audit, le seuil enregistré, l'attestation conservée comme l'étalon au regard duquel la conformité est jugée, et restreindre M à ce registre l'empêche d'enfler en omnivore. Dans le vocabulaire des travaux antérieurs de cet institut, le résidu est le domaine de la *reconstructibilité* : ce qui ne peut être reconstruit après coup, parce que rien n'a été consigné à partir de quoi reconstruire.

La mémoire porte aussi sur la réversibilité, quoique pas de la manière simple qu'un premier examen suggère. La mémoire ne rend pas, par sa seule existence, le retour coûteux ; une mémoire bien construite peut le rendre bon marché, par le versionnage, l'historique de décision, le rollback, un enregistrement conservé de l'ancien seuil. Ce qui rend le retour coûteux, c'est une structure particulière de mémoire : l'état dégradé devient la référence par défaut, l'état antérieur perd son statut, la restauration exige une nouvelle autorisation, et la charge de preuve entre l'abaissement et le relèvement n'est pas

symétrique. L'irréversibilité n'est pas produite par la mémoire en tant que telle. Elle est produite par l'asymétrie de ré-inscription et de restauration qu'une mémoire donnée encode. Nommer cette asymétrie nomme le remède, et le remède est plus qu'instrumenter les absences. Une mémoire construite contre le cliquet versionnerait ses politiques, conserverait le niveau antérieur au lieu de l'écraser, exigerait une justification signée pour chaque dégradation, ferait expirer les exceptions automatiquement au lieu de les laisser se déposer dans la ligne de base, tiendrait la charge de preuve symétrique, et garderait un rollback à portée. Le but n'est pas d'enregistrer davantage. C'est de rendre les dégradations non auto-perpétuantes, afin qu'un seuil abaissé décroisse vers l'étalon à moins d'être activement renouvelé, plutôt que de s'y durcir par défaut.

5. Deux régimes de feedback

La dynamique peut maintenant être énoncée avec l'économie que les primitives autorisent, et le mot *attracteur* doit être écarté ; il promet un espace d'états, des équations de transition et des conditions de stabilité que l'argument ne fournit pas. Appelons-les *régimes de feedback*.

Le régime bruyant est *en quête de correction*, et le terme est choisi avec soin : il n'est pas *auto-correcteur*, ce qui revendiquerait trop. Le signal crée une pression de correction ; il ne garantit pas que la correction aboutisse. Le système peut être immuable, la cause externe, les utilisateurs peuvent contourner sans améliorer, les gestionnaires peuvent ignorer les contournements. Ce que le régime bruyant garantit, c'est un signal qui pèse sur la boucle, non une réparation. Même ainsi, la pression est réelle et pointe vers la correction, ce qui est précisément ce dont le régime silencieux manque.

Le régime silencieux n'est pas en quête de correction, et l'assertion exacte doit être prudente. L'absence de correction ne fait pas croître par elle-même l'erreur silencieuse. Elle la fait persister. Trois états restent distincts, et ils séparent deux dynamiques différentes. La *persistance silencieuse* est l'erreur demeurant à politique constante : invisible, stationnaire, une charge latente. L'*accumulation normative silencieuse* est la dynamique de politique, des déclassements successifs, chacun abaissant la barre dont part la décision suivante, déplaçant Π elle-même. L'*amplification silencieuse de l'exposition* est la dynamique de volume, davantage de cas envoyés à travers la politique dégradée sans autre changement de Π . Le premier état est inerte. Le cliquet vit dans le deuxième et, par lui, dans le troisième, et les deux ne doivent pas être confondus, car l'un est une dérive de la règle et l'autre une dérive de ce à quoi la règle s'applique.

La migration du régime bruyant vers le régime silencieux est ce qu'un titre antérieur nommait le *cliquet*. ***Rester dans le régime bruyant est coûteux, puisque rester signifie payer la friction encore et encore ; se déplacer vers le régime silencieux est bon marché, celui-ci étant tranquille et non surveillé.*** Le système ne manque pas de

corriger l'erreur silencieuse par négligence. Il est tiré, par la forme de son feedback et l'asymétrie de son registre, hors de la région qu'il peut corriger et dans la région qu'il ne peut voir. L'irréversibilité de cette migration est une hystérésis au sens strict de dépendance au chemin, non l'hystérésis magnétique de la physique : l'état après une séquence de déclassements n'est pas une fonction des seules pressions courantes mais du registre que ces déclassements ont consigné, et ce registre ne s'efface pas quand la pression s'efface.

6. Une prédiction, sa mesure, et le programme qui la testerait

Une théorie devient intéressante quand elle prédit quelque chose que le cadre en place ne peut prédire, assez nettement pour pouvoir avoir tort. Prenez deux portes de fiabilité aux performances initiales identiques, précision identique, rappel identique. La lecture classique les dit équivalentes. La loi prédit que si elles diffèrent par la force contraignante que leur architecture assigne aux deux classes d'erreur, elles dériveront à des rythmes différents et dans des mesures différentes, nonobstant la performance initiale.

L'écart mesuré appelle sa propre définition, puisqu'un passage de la classe un à la classe deux n'est pas un passage de la classe deux à la classe trois, et qu'un décompte non pondéré les traiterait à l'identique. Soit l'*écart de promotion* au temps t une somme pondérée sur les ports, $G(t) = \sum w_i d(R_i\text{cible}, R_i\text{réel})$, où d est une distance définie par le domaine entre classes de charge de preuve, non une soustraction. Les classes sont ordinales, les poids seuls ne fournissent pas la métrique manquante, et d doit se construire à partir des coûts ou des exigences que le domaine attache à chaque classe, ou se remplacer par un indicateur ordinal tel que le décompte pondéré des transitions de dégradation. La vélocité de dérive est $v = \Delta G / \Delta t$; la distance de dérive à un horizon H est $D_H = G(H) - G(0)$. La formalisation est illustrative, et son service est de fixer ce qui serait enregistré.

Ceci se lit mieux comme un programme expérimental que comme une expérience unique, et trois difficultés l'empêchent d'être propre. Échantillonner les décisions ALLOW change plus que la force contraignante : cela change le coût d'audit, la charge humaine, la disponibilité des étiquettes, la fréquence de correction, possiblement le modèle lui-même, de sorte que l'intervention n'isole pas parfaitement l'architecture du feedback. Les conditions réclament randomisation ou simulation, puisque des opérateurs qui se savent observés s'adaptent à l'être. Et l'unité d'analyse doit être fixée, porte ou port ou décision ou période ou équipe ou type de risque, puisque l'effet peut apparaître à un grain et disparaître à un autre. Rien de tout cela ne rend la prédiction intestable. Cela fait du test un programme conçu, ce qui est la description honnête de ce qu'exige une assertion sur la dynamique institutionnelle.

L'unique illustration disponible dans l'intervalle doit être cadrée pour ce qu'elle est et rien de plus. Une instance de fiabilité pour la prédiction toxicologique, construite sur une featurisation moléculaire qui encode les atomes par un numéro atomique en one-hot, porte un angle mort structurel pour les complexes de coordination métalliques, dont le comportement dépend d'états d'oxydation et d'une chimie des ligands que la featurisation ne représente pas. Un faux ALLOW confiant sur un tel composé est une erreur silencieuse au sens exact employé ici : accepté à l'exécution, force contraignante proche de zéro, pas d'artefact, pas d'entrée au registre. Cette instance a été mesurée ailleurs². Cela montre que la classe silencieuse n'est pas une commodité théorique ; elle a une instance concrète avec une cause concrète. Une instance prouve l'existence, non la généralité, raison pour laquelle le programme ci-dessus est dû plutôt que présumé.

7. Ce que cela déplace, et quelle force est isolée

Les comparaisons se lisent mieux comme un déplacement de priorité causale que comme la découverte d'un terrain vierge. Ces corpus ne sont pas aveugles à la sélectivité du feedback. Vaughan (1996) voit le non-événement qui n'alarme pas ; la sûreté organisationnelle a bâti tout un appareillage autour de ce qui est rapporté et ignoré. Ce que ce cadre fait, c'est ériger cette sélectivité en objet primaire plutôt qu'en un facteur parmi d'autres, et la lier à l'architecture de la porte et à la transformation de la politique. Sous ce cadrage, chaque distinction est un changement d'accent et de mécanisme, non une prétention d'absence. Levitt et March décrivent l'apprentissage à partir d'un signal reçu ; ici l'objet est la condition sous laquelle le signal existe. De Holan et Phillips décrivent la perte d'un savoir acquis ; ici il s'agit d'un savoir jamais inscrit. Le biais d'automatisation, d'après Parasuraman et Manzey (2010) et Mosier et ses collègues (1998), siège dans la cognition de l'opérateur ; ici le niveau est la politique institutionnelle. La loi de Goodhart, dans la formulation de Strathern (1997), décrit une mesure corrompue par le fait de devenir une cible ; ici c'est le port qui se dégrade sous un champ que la mesure ne capte pas.

La littérature sur les quasi-accidents appelle un placement exact. Un quasi-accident laisse habituellement bien une trace ; l'organisation la pondère simplement faiblement, car aucune conséquence sérieuse n'a suivi. Le problème là n'est pas un signal absent mais un signal faiblement contraignant, ce qui place les quasi-accidents non à l'extrême silencieux mais à l'intérieur d'un continuum que la théorie devrait énoncer clairement : un signal peut être absent, présent mais faiblement contraignant, ou fortement contraignant, et la dérive est gouvernée par l'endroit de ce continuum où tombe le

² J. Vétillard, « Un domaine d'applicabilité mesure une proximité, pas une capacité », Twingital Institute, 15 juillet 2026. L'audit y établit, sur un service de prédiction toxicologique, qu'une featurization purement topologique écrase l'information distinguant des composés proches, et que le port de validité certifie une conformité sans référent : instance mesurée du faux ALLOW silencieux et de l'écart de promotion (Article VI).

feedback d'une erreur. L'erreur silencieuse est le cas limite où la force contraignante atteint zéro ; le quasi-accident est le milieu instructif.

La littérature clinique sur la fatigue d'alerte relève d'ici comme instance plutôt que comme rivale. Les taux de contournement rapportés pour les alertes médicamenteuses vont d'environ quarante-six à quatre-vingt-seize pour cent, et les institutions répondent en supprimant des alertes et en reconfigurant des seuils. Lue à travers la lentille classique, c'est de la réduction de bruit ; lue à travers la loi, c'est le cliquet observé sur le terrain sous un autre nom, la suppression du signal bruyant étant le mécanisme par lequel la politique migre vers le régime tranquille. Les données sont fortes comme illustration et faibles comme preuve, puisqu'elles ont été recueillies pour étudier la fatigue et non la dérive.

L'isolement de la force contraignante demande une défense exacte. Appeler B structurel et le reste perturbations confondrait deux distinctions : une régulation peut être aussi structurelle qu'une architecture, un budget aussi durablement inscrit qu'un port. La distinction honnête n'est pas structurel contre perturbation mais endogène contre exogène à l'architecture du feedback, sur la frontière fixée en Section 2. B_A est endogène, une propriété de la façon dont la porte transforme les erreurs en signaux. Régulation, budget, culture, leadership sont exogènes, et peuvent expliquer davantage de la variance dans un cas donné que ne le fait B. B est isolé non parce qu'il domine mais parce qu'il est la seule force que l'architecture elle-même produit, et donc la seule qu'un architecte puisse changer.

8. La configuration immunitaire, le périmètre, et les limites

Le mécanisme a une condition d'annulation, et la nommer fait plus de travail que ne le pourrait aucune défense du mécanisme. Que la promotion soit signée par une partie indépendante, les décisions ALLOW échantillonnées systématiquement, le point d'ajustement déplacé hors de l'exécution vers un cycle d'audit distinct. Alors les forces de rappel exogènes rencontrent la force endogène, le coût de la dégradation s'élève vers le coût de la promotion, les absences sont instrumentées et la mémoire cesse d'être biaisée, et la dérive ne s'amorce pas. C'est la condition d'annulation et une hypothèse de conception, non une réfutation propre. Un système unique qui satisfait la configuration et dérive malgré tout prouve peu, puisque la dérive peut venir d'un auditeur capturé, d'un registre falsifié, d'une revue sous-échantillonnée, d'un choc budgétaire, de n'importe laquelle des forces exogènes que la loi admet. Réfuter la loi requiert une comparaison contrôlée où ces forces sont tenues immobiles et où seule l'architecture du feedback varie.

Les domaines nommés plus haut comme sites d'extension méritent une description plus nette, ce qui renforce l'argument. Le signalement des quasi-accidents dans l'aviation, l'analyse des précurseurs dans le nucléaire, le signalement en culture juste des secteurs à haute fiabilité ne sont pas un terrain vierge attendant la théorie ; ils sont le mécanisme déjà pris pour cible par l'ingénierie, chacun un appareillage destiné à donner un signal contraignant à l'erreur qui, sans lui, passerait silencieuse. La comparaison doit séparer trois quantités, sans quoi elle ne prouve rien : le dispositif nominal, le taux réel d'inscription qu'il atteint, et la force contraignante réelle qu'il porte dans la décision. Un système de quasi-accidents à faible taux d'inscription réel n'est une configuration immunitaire que de nom. Le corollaire falsifiable survit à la réserve et en sort plus tranchant : les secteurs qui atteignent une inscription réelle élevée et une force contraignante réelle élevée devraient dériver plus lentement que les secteurs dont les instruments sont nominaux.

Sur le périmètre, la position tient : le principe énoncé en général, la loi démontrée seulement là où ses quantités peuvent être mesurées, l'extension offerte comme conjecture. Une tentation est mise à l'écart par cloison. Il est tentant de conclure que les institutions convergent non vers la vérité mais vers ce qui contraint le plus fort, et d'énoncer cela comme une loi. C'est une formule doctrinale, gardée comme telle, utile comme compression et disqualifiée comme autorité ; la loi est conditionnelle et ne vaut que là où une boucle de décision locale gouverne l'ajustement de la politique. Les limites sont portées sur la page : la théorie est plus développée que sa preuve, laquelle repose sur une instance et un corpus de données cliniques réaffecté ; le programme central est spécifié mais non exécuté ; le rapport de coûts est argumenté mais non quantifié ; la force contraignante reçoit un proxy, le coût normalisé de suppression, lui-même impur et non validé.

9. Conclusion

La porte de fiabilité a d'abord été offerte comme un moyen d'encoder un compromis. Elle encode aussi autre chose, sans que personne décide qu'elle le doive. En rendant une erreur bruyante et l'autre silencieuse, elle fixe la force contraignante de leurs signaux, et par cette force elle fixe quelles erreurs atteignent le registre et lesquelles n'y parviennent jamais. La politique se déplace alors vers les erreurs qu'elle ne peut ignorer et loin des erreurs qu'elle ne peut voir, et le registre formel, ne retenant que ce qui a laissé une trace, transforme le mouvement en règle permanente par la seule asymétrie qui compte : il est moins coûteux d'inscrire l'état dégradé que de l'inscrire en retour.

Le mécanisme qui a donné à cet article son premier titre est la moindre part de ce qu'il revendique désormais. La part plus grande est conditionnelle et, parce que conditionnelle, testable : à distributions d'erreurs initiales égales, les politiques institutionnelles divergent lorsque leurs architectures de feedback imposent des forces

contraignantes différentes à la boucle de décision locale. Une institution n'enregistre pas ce qui lui arrive. Elle enregistre ce qui laisse une trace, et quelque chose décide à l'avance ce qui a le droit d'en laisser une. Cette chose qui décide est l'architecture du feedback. Elle vaut d'être nommée, car une dérive située dans l'architecture peut être contrée par l'ingénierie, plutôt que d'être traitée par la seule culture. Une culture de sûreté peut changer ce qui est rapporté, attribué et inscrit, et dans cette mesure elle change aussi l'architecture. La thèse est seulement que l'architecture est le lieu où la dérive s'écrit, et donc le lieu où on peut la contrer.