Consensus, Distillation, and Trust: On the Mathematics of Agreement in Machines, Networks, and People

A. C. Applebee[a] and L. N. Combe[a]

Preprint, not peer reviewed • Version 1.0-preprint • 2026-10-14 • Licence CC BY 4.0

Abstract

How do distributed agents, whether miners, neurons or people, come to agree without a central authority? This paper treats three cases with mathematics at three stated levels of rigour. Part I gives the standard treatment of Bitcoin's proof-of-work: exponential inter-block times, the double-spend probability as a regularised incomplete beta function, and the optimality of honest mining in the absence of difficulty adjustment. Part II develops a formal analogy between knowledge distillation in neural networks and human behavioural learning, with implications for criminology's rational actor model: not a dissolution of responsibility but a richer account of how behavioural patterns propagate. Part III analyses social consensus for identity and vote counting, a web-of-trust design in which each person must be vouched for in person by others, as a case study in graph-based Sybil resistance whose guarantees are conditional on measurable costs. We add two reproducible checks. On 4,032 real Bitcoin blocks (heights 840000 to 844031), inter-block times fit the exponential law of Theorem 1 (coefficient of variation 0.994; Kolmogorov-Smirnov $p = .50$; hourly counts Poisson, dispersion 1.05). Numerical checks confirm the double-spend formula against a simulation of one million races per cell, confirm the corollary's asymptotics, and show the high-temperature distillation limit holds for mean-centred logits, the condition Hinton and colleagues assume. All three systems reach agreement through costly signals; they differ in how precisely the cost can be measured.

Keywords: blockchain | proof of work | knowledge distillation | social learning theory | Sybil resistance | Poisson processes | rational actor | criminology

Significance. The same problem, agreement among agents who cannot all be trusted, recurs in money, in learning and in identity. Placing the mathematics side by side, at honestly stated levels of rigour, shows criminology a formal model of learned behaviour and shows identity designers exactly which number their security rests on.

Introduction

Bitcoin, introduced by Nakamoto in 2008, solved the problem of decentralised monetary consensus: how can a network of mutually distrusting computers agree on the state of a shared ledger (Nakamoto, 2008)? The answer rests on cryptographic hash functions, Poisson processes and game-theoretic incentives. A child learning to regulate emotion solves a different-looking problem, converging on the behaviour of the people around it with no rulebook. A community counting its members, so that each person gets one vote or one share, faces a third: how to agree who is real without a central registry.

This paper treats all three as instances of distributed agreement. We are explicit throughout about where the mathematics is rigorous (Part I), where it is a structural analogy that generates testable predictions (Part II), and where it is sound but conditional on parameters that are hard to measure (Part III). Part I is a rigorous benchmark against which the less formal arguments can be judged.

Part I. Computational consensus: Bitcoin

Cryptographic primitives. A cryptographic hash function $H : \{0,1\}^* \to \{0,1\}^{256}$ is preimage resistant (given y, finding x with $H(x) = y$ is infeasible), collision resistant (finding $x \neq x'$ with $H(x) = H(x')$ is infeasible) and pseudo-random (changing one input bit flips each output bit with probability 1/2). Bitcoin uses SHA-256 for block hashing, RIPEMD-160 for addresses and ECDSA over secp256k1 for signatures. Each block header contains the hash of the previous block, a Merkle root committing to its n transactions with $O(\log n)$ membership proofs, a timestamp, the difficulty target and a nonce.

The mining process as a Poisson model. Let a miner control a fraction $p \in (0, 1]$ of the network's hashrate, with blocks found network-wide on average every $\tau_0 = 10$ minutes.

Theorem 1 (mining time). The time T between successive blocks found by this miner is exponential with rate $\alpha = p/\tau_0$: $f_T(t) = \alpha e^{-\alpha t}$, $t \geq 0$.

Proof. Pseudo-randomness makes each hash attempt independent. Let $P(t)$ be the probability of no valid block by time t. Independence over disjoint intervals gives $P(t + s) = P(t)P(s)$, whose only continuous solution is $P(t) = e^{-\alpha t}$; $E[T] = \tau_0/p$ gives $\alpha = p/\tau_0$. □

The sum of n inter-block times is Gamma(n, α) and the count $N(t)$ of blocks by time t is Poisson with mean αt.

The double-spend problem. An attacker with hashrate fraction q (honest fraction $p = 1-q$) tries to replace a confirmed transaction by secretly mining an alternative chain.

Theorem 2 (double-spend probability; Grunspan & Pérez-Marco). After z confirmations, an attacker with $q < 1/2$ succeeds with probability

$$P(z) = I_{4pq}(z, 1/2),$$

where $I_x(a, b)$ is the regularised incomplete beta function (Grunspan & Pérez-Marco, 2018a).

Corollary 3 (exponential decay). With $s = 4pq < 1$, $P(z) \sim s^z / \sqrt{\pi(1-s)z}$ as $z \to \infty$.

Table 1. Double-spend probabilities as given in the February 2026 draft. Recomputations under Theorem 2 and under Nakamoto's approximation are in Table 2.

z	$q = 0.10$	$q = 0.20$	$q = 0.30$	$q = 0.40$
1	0.2045	0.3781	0.5310	0.6696
2	0.0509	0.1531	0.2937	0.4560
3	0.0134	0.0641	0.1653	0.3138
4	0.0036	0.0272	0.0938	0.2175
5	0.0010	0.0117	0.0536	0.1514
6	0.0003	0.0050	0.0308	0.1058

Table 2. Recomputation: Theorem 2 (exact) / Nakamoto's 2008 Poisson approximation. A simulation of one million races per cell agrees with the exact column to within 0.002 in every cell. Source: `studies/01-theorem-checks`.

z	$q = 0.10$	$q = 0.20$	$q = 0.30$	$q = 0.40$
1	0.2000 / 0.2046	0.4000 / 0.4159	0.6000 / 0.6277	0.8000 / 0.8289
2	0.0560 / 0.0510	0.2080 / 0.2039	0.4320 / 0.4457	0.7040 / 0.7364
3	0.0171 / 0.0132	0.1158 / 0.1032	0.3262 / 0.3246	0.6349 / 0.6642
6	0.0006 / 0.0002	0.0233 / 0.0143	0.1564 / 0.1321	0.4930 / 0.5040

The draft's $q = 0.10$ column reproduces Nakamoto's published approximation (Nakamoto, 2008). Its columns for $q = 0.20$ to 0.40 come from a calculation we could not reconstruct from either formula and run lower than both; Table 2 gives both recomputations so readers can see the range. The qualitative result stands in every version: the attacker's chance falls geometrically with each confirmation, and falls slowly when the attacker is strong. Corollary 3 is accurate: the ratio of exact to asymptotic value is 0.93 at $z = 5$ and 0.99 at $z = 40$ ($q = 0.10$).

Protocol stability. **Theorem 4 (honest mining is optimal; Grunspan & Pérez-Marco).** Without difficulty adjustment, publishing every block on discovery maximises revenue (Grunspan & Pérez-Marco, 2018b).

Proof sketch. Define the revenue ratio $\Gamma = E[R]/E[\tau]$ per cycle. $M(t) = N(t) - \alpha t$ is a martingale; by Doob's optional stopping theorem $E[N(\tau)] = \alpha E[\tau]$ for any stopping time, so $\Gamma \le pb/\tau_0 = \Gamma_H$, with equality only for honest mining. □

Remark 5 (selfish mining). Withholding blocks becomes profitable when $q > (1-\gamma)/(3-2\gamma)$, where γ is the share of honest miners who build on the attacker's block when two blocks tie; for $\gamma = 0.5$ the threshold is $q > 0.25$ (Eyal & Sirer, 2018). Bitcoin adjusts difficulty every 2,016 blocks, which selfish mining can exploit, so the gap between the idealised and practical security models remains open.

Part II. Distillation as a theory of learned behaviour

From neural networks to human learning. Knowledge distillation trains a simple "student" network to reproduce the outputs of a complex "teacher" (Hinton et al., 2015). The student never sees the teacher's internal reasoning, only its outputs, given as soft probability distributions; yet it learns to approximate the teacher's competence closely. We argue this is more than a metaphor for human behavioural learning. A child learning emotional regulation from a caregiver receives no explicit rules. The child observes the caregiver's responses to situations and, through repeated exposure and consequence, learns to approximate them. The knowledge is never transmitted; it is distilled from observed behaviour.

Formal framework. **Definition 6 (knowledge distillation).** Let $f_T, f_S : \mathcal{X} \to \Delta^{C-1}$ be teacher and student models with logits z_T, z_S. Distillation minimises

$$\mathcal{L}_{KD} = (1-\lambda)\,\mathcal{L}_{CE}(f_S, y) + \lambda T^2\,\mathrm{KL}\big(\sigma(z_T/T)\,\|\,\sigma(z_S/T)\big),$$

where σ is the softmax, $T > 1$ the temperature, y the true label and $\lambda \in [0, 1]$.

The temperature controls how much "dark knowledge" passes. At $T = 1$ the student sees almost only the teacher's top answer. As T rises, the softer distribution reveals the teacher's uncertainty: which wrong answers are almost right, which classes it confuses.

Proposition 7 (dark knowledge transfer). As $T \to \infty$, with logits centred to zero mean,

$$T^2 \,\mathrm{KL}\big(\sigma(z_T/T) \,\|\, \sigma(z_S/T)\big) \to \frac{1}{2C} \sum_{c=1}^{C} (z_{T,c} - z_{S,c})^2.$$

The student learns not just what the teacher predicts but the relative structure of its uncertainty.

The centring condition matters because the softmax ignores any constant added to all logits; Hinton and colleagues assume zero-mean logits for the same reason (Hinton et al., 2015). Our numerical check (2,000 random pairs, 10 classes) finds the median relative error of the approximation falls from 7.3% at $T = 5$ to 0.4% at $T = 100$ for centred logits, while uncentred differences do not converge.

Graph-structured distillation. Social data have graph structure. A graph neural network computes $h_v^{(\ell+1)} = \phi\big(h_v^{(\ell)}, \bigoplus_{u \in N(v)} \psi(h_v^{(\ell)}, h_u^{(\ell)}$ aggregating each node's neighbourhood; distillation can extract simple explanations of which neighbourhood structures drive a graph network's predictions (Pereira et al., 2023). Here the analogy to social learning becomes precise: each node's representation is shaped by its neighbours through repeated message passing, as a person's repertoire is shaped by their social environment through repeated interaction.

The mapping.

Table 3. Structural correspondence between knowledge distillation and human behavioural learning. This is not a proof; it is a correspondence we argue is productive.

Component	Distillation	Human learning
Teacher	Complex model f_T	Caregiver, social environment
Student	Simple model f_S	Developing person
Input	Data point x	Situation, stimulus
Output	Soft distribution $\sigma(z/T)$	Behavioural response
Dark knowledge	Inter-class uncertainty	Implicit emotional cues
Temperature	Softness of the signal	Emotional expressiveness
Loss	$\mathcal{L}_{KD}$	Reinforcement, consequence
Graph	$N(v)$ in a GNN	Social network

What this explains that rational choice does not. The rational actor model in criminology holds that offending results from utility-maximising deliberation: a person weighs expected benefits against expected costs and chooses crime when the payoff exceeds the risk (Becker, 1968; Cornish & Clarke, 1986). The model has been productive but struggles with three well-documented patterns. Much violent and impulsive offending shows no sign of cost-benefit calculation. Offending runs in families and communities at rates rational imitation would not predict. And the same person behaves very differently across social settings, which stable preferences do not explain.

The distillation account offers a complementary explanation. If people learn responses by approximating the soft outputs of their social environment (not its explicit rules but its implicit distribution over responses), then *impulsive behaviour* corresponds to a student trained at high temperature, which absorbed the full uncertainty of its environment, volatile and extreme responses included; *intergenerational transmission* corresponds to iterated distillation, each generation distilled from the last, with compounding approximation error, as in repeated student-to-student distillation (Furlanello et al., 2018); and *context sensitivity* corresponds to domain shift, where a model trained on one distribution behaves unpredictably on another.

Relationship to existing theories. These observations are not new. Social learning theory established that behaviour is learned through observation and reinforcement (Bandura, 1977); differential association theory, that criminal behaviour is learned in intimate groups (Sutherland, 1947); strain theory, that social structure shapes behavioural outcomes (Merton, 1938). Distillation does not replace them. It adds a formal mechanism, the KL-divergence loss, a precise description of "approximating observed behaviour" that could in principle be tested on behavioural data; the dark-knowledge insight, that what passes from caregiver to child is not "anger is the right response" but a whole distribution ("anger is likely, withdrawal possible, calm unlikely"); and graph-structured propagation, since social learning happens in networks, not pairs.

On responsibility. If behaviour is learned through an unconscious, distillation-like process, does anyone bear responsibility for their actions? We think the answer is plainly yes, and that the framework shows why. A network trained by distillation is responsible for its outputs in the engineering sense: we evaluate it, deploy it or not, and retrain it if it fails. That it learned from a teacher does not absolve the deployed model; the teacher is the causal history, the current weights are the present reality. The same holds for people. A pattern acquired decades ago is now the person's own; it runs through their nervous system and affects other people. The origin of a problem does not decide its ownership.

What changes is the response. Criminal justice should focus less on punishing rational choices that may not have occurred and more on retraining: new data, new environments, new feedback that update the learned distribution. In machine learning a badly behaved model is corrected by fine-tuning; the analogue of rehabilitation is

$$\mathcal{L}_{\text{rehab}} = \mathrm{KL}(\pi_{\text{prosocial}} \,\|\, \pi_{\text{current}}),$$

minimised through structured environments, modelling of alternative responses and reinforcement. Fine-tuning works precisely because the model owns its current weights; responsibility and learned behaviour are complementary, not opposed. Evidence-based rehabilitation already works this way.

Part III. Social consensus for identity and vote counting

Scope. Part I established rigorous results; Part II a suggestive analogy. This part sits between: a framework that is internally consistent but whose security depends on an empirical parameter, the cost of social fraud, that is hard to measure. We focus narrowly on counting unique people for fair distribution, analogous to counting votes, using as a case study a web-of-trust design in which every new member must be vouched for in person by existing members. We do not claim security comparable to Bitcoin's; the threat models differ.

The Sybil problem. A Sybil attack is the creation of many fake identities by one entity to gain disproportionate influence. Without a trusted central authority, Sybil attacks cannot be prevented in a purely open network (Douceur, 2002); every practical defence imposes a cost on creating identities. Bitcoin imposes a computational cost; social-network defences impose a social one (Yu et al., 2006).

Web-of-trust verification. **Definition 8 (verification graph).** $G = (V, E)$ is directed, V the verified identities, $(u, v) \in E$ if u vouched for v; every vertex has in-degree $d^-(v) \geq k$ (the design studied uses $k = 3$).

Definition 9 (vouch cost). $c_{\text{vouch}} = c_{\text{physical}} + P[\text{detection}] \cdot c_{\text{penalty}}$, where c_{physical} is the cost of an in-person verification and c_{penalty} the expected loss on detection (reduced trust, ejection).

Proposition 10 (linear Sybil cost). Creating n Sybil identities needs at least nk fraudulent vouches; with independent vouch costs, $C(n) \geq n \cdot k \cdot c_{\text{vouch}}$.

This is an argument, not a theorem in Part I's sense: it depends on estimating c_{vouch}, on vouchers being independent (collusion breaks this) and on detection being non-trivial.

Definition 11 (collusion indicator). For the vouchers S of a candidate, $I(S) = |E(S)|/\binom{|S|}{2}$, where $E(S)$ is the edge set of the subgraph S induces. High values flag possible collusion. It is a heuristic: attackers can keep $I(S)$ low by using vouchers who are not connected. It raises the cost of attack; it does not remove it.

Theorem 12 (equal per-capita distribution). If the verification graph contains no Sybil identities, the rule $r_i(t) = R(t)/|V(t)|$ allocates resource $R(t)$ equally. *Proof.* Immediate. The content of the theorem is that the mechanism is correct given its assumption; the security question is how well the web of trust enforces the assumption. The system's fairness is exactly as good as its Sybil resistance.

Anchoring to Bitcoin. The design can inherit Bitcoin's security for data integrity by committing a Merkle Mountain Range root R_e of all identity records to a Bitcoin transaction each epoch e. After z confirmations, reverting R_e requires a Bitcoin double-spend, with probability $I_{4pq}(z, 1/2)$ by Theorem 2. This is the one place Bitcoin-grade security genuinely transfers: not to identity verification itself, but to the immutability of committed records.

Two checks of our own

Study 1: numerical verification. We recomputed every quantitative claim in Parts I and II (`studies/01-theorem-checks`): Theorem 2 against a Monte Carlo race (one million races per cell, fixed seed), Nakamoto's approximation, the asymptotics of Corollary 3, Proposition 7 on 2,000 random logit pairs at four temperatures, and the selfish-mining threshold. Results are reported in Part I and Part II above.

Study 2: real block times. Theorem 1 predicts exponential gaps between blocks. Real timestamps are set by miners and are noisy, so the prediction is not trivially true. We took the 4,032 blocks of two difficulty periods, heights 840000 to 844031 (April to May 2024), from a public block explorer (mempool.space, 2026).

Fourteen gaps were negative (timestamps out of order) and were set aside. The remaining 4,017 had a mean of 620 seconds against the 600-second target, a median of 440 seconds (the exponential predicts 430), and a coefficient of variation of 0.994 (the exponential predicts 1). A Kolmogorov-Smirnov test does not reject the exponential (D = 0.013, p = .50). Counted by the hour, blocks are Poisson: mean 5.82, variance 6.14 (dispersion index 1.05), with no significant departure from Poisson frequencies (χ^2 = 10.4, p = .40). The cleanest model in the paper holds on live data.

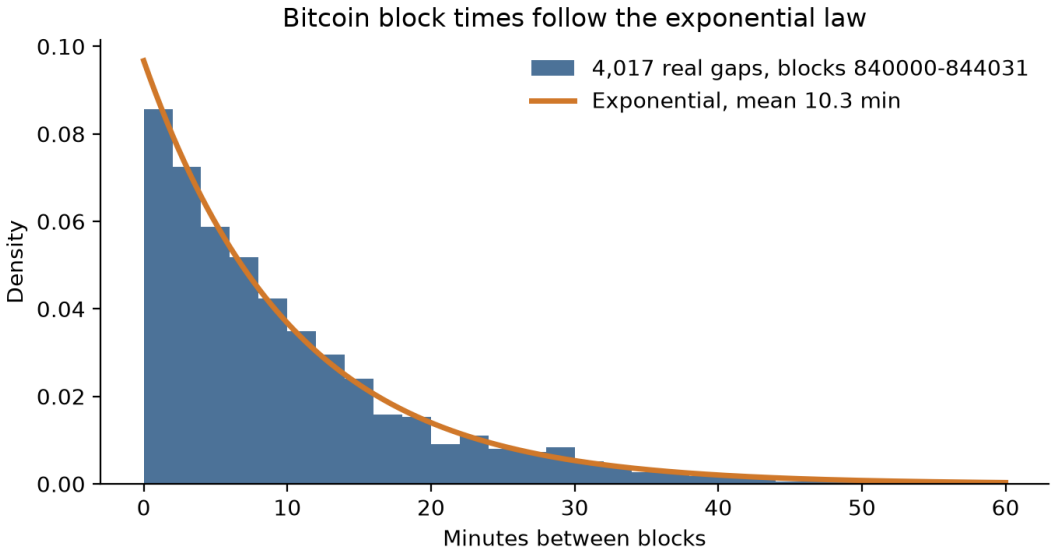


Fig. 1. Gaps between 4,017 consecutive Bitcoin blocks (navy) against the exponential distribution with the same mean (orange).

Synthesis

Table 4. Three consensus mechanisms and their properties.

	Bitcoin	Distillation	Web of trust
Agents	Miners	Neurons, people	Verified people
Agreement on	Transaction order	Behavioural distribution	Identity uniqueness
Cost of deviating	Wasted hashrate	Prediction error, consequences	Social cost
Guarantee	Strong (Theorem 2)	Bounded (Proposition 7)	Conditional (Proposition 10)
Open problem	Selfish-mining gap	Causal identification	Measuring c_{vouch}

All three reach agreement through costly signals. In Bitcoin the signal is a valid proof of work, costly in energy and hardware. In distillation and human learning it is behavioural consistency with one's environment, costly because deviation brings consequences: prediction error for networks, social punishment for people. In a web of trust it is a vouch, costly because the voucher stakes their own standing. In each case the cost is what stops the system being trivially exploited; the systems differ in how precisely the cost can be measured, and so in how rigorous their guarantees can be.

Open problems

1. *Measuring social cost.* Can c_{vouch} be measured, for example in mechanism-design experiments? Until then web-of-trust security stays qualitative.
2. *Testing the distillation analogy.* "High-temperature" environments (emotionally volatile, inconsistent caregiving) should produce broader behavioural variance. Do longitudinal data agree?
3. *Generation loss.* Repeated student-to-student distillation degrades; does intergenerational transmission show an analogous loss, and what does that imply for when to intervene?
4. *The selfish-mining gap* between idealised and real Bitcoin security.
5. *Graph-based social learning.* Can message-passing models be calibrated against real social influence data?

Limitations

Part II is an analogy with testable consequences, not a demonstration; children are not gradient descent. Part III's guarantees are conditional on c_{vouch}, voucher independence and detection, none yet measured. Study 2 covers two difficulty periods only, and miner-set timestamps add noise that slightly widens the gap distribution. The draft's Table 1 columns for $q \geq 0.20$ could not be reconstructed; we print both recomputations beside it rather than replace it.

Conclusion

We set out three views of distributed consensus at three levels of rigour: Bitcoin's proof-of-work, where exact probabilities follow from the incomplete beta function and honest mining's optimality from martingale theory, and where real block times obey the

model; knowledge distillation as a model of human behavioural learning, where the formal machinery is precise and the analogy productive, but causal claims need data; and social consensus for identity, where the framework is sound and the key parameter resists measurement. Placing them side by side does not pretend they are equally rigorous. It shows the same structural problem recurring across domains, and tools from one domain lighting up another. For criminology in particular, distillation offers a formal model of learning behaviour from observation that specifies a mechanism, makes predictions and connects to a large body of theory.

Consensus is not agreement imposed from above. It is agreement that emerges from cost, structure and repeated interaction, whether the agents are machines, neurons or people.

Data and code availability

`reproduce.sh` reruns both studies and checks them against `expected/`: `studies/01-theorem-checks` (theorem table, formulas, simulations; the February 2026 validation script is kept as `src/original_run.py`) and `studies/02-block-times` (4,032 block timestamps fetched once by `src/fetch.py` and committed). The statistical appendix (`appendix/statistical-appendix.pdf`) sets out each test, its assumptions, power, a sceptic's question and the answer. Repository: https://github.com/vexanonus/consensus-distillation.

Appendix: statistics

Monte Carlo race (Study 1). *Why:* an independent referee for Theorem 2 that uses no special functions. The honest network finds z blocks; the attacker's blocks in that time are negative binomial; from a deficit d the attacker catches up with probability $(q/p)^d$ (gambler's ruin). One million races per cell give a standard error under 0.0005. *Sceptic:* "Different race models give different numbers." *Answer:* yes; that is why Table 2 shows Nakamoto's approximation beside the exact formula, and why the simulation follows the exact formula's model.

Kolmogorov-Smirnov test (Study 2). *Why:* compares the whole distribution of gaps with the exponential, not only the mean. *Assumptions:* independent gaps, a continuous distribution; the rate is estimated from the data, which makes the test conservative (a Lilliefors correction would lower *p* slightly). *Power:* with 4,017 gaps the test detects departures of about 0.02 in the cumulative distribution; the observed *D* is 0.013. *Sceptic:* "Hashrate changes within an epoch." *Answer:* it does, which would make gaps over-dispersed; the dispersion index of 1.05 shows little of it in this window.

Chi-square goodness of fit for hourly counts. Cells with expected counts below five pooled out; one degree of freedom removed for the estimated mean.

Proposition 7 check. Median relative error over 2,000 random logit pairs (teacher and student logits normal with SD 2, student shifted by a random constant) at T = 1, 5, 20, 100.

Scripts: `studies/01-theorem-checks/src/run.py`, `studies/02-block-times/src/run.py`, `src/figure.py`.

Related papers: Applebee, A. C., & Combe, L. N. (2026). *Learning distillation*; Applebee, A. C., & Combe, L. N. (2026). *The language proof*.

References

Bandura, A. (1977). *Social learning theory*. Prentice Hall.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, *76*(2), 169–217. https://doi.org/10.1086/259394

Cornish, D. B., & Clarke, R. V. (1986). *The reasoning criminal: Rational choice perspectives on offending*. Springer. https://doi.org/10.1007/978-1-4613-8625-4

Douceur, J. R. (2002). The Sybil attack. In *Lecture notes in computer science* (pp. 251–260). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45748-8_24

Eyal, I., & Sirer, E. G. (2018). Majority is not enough: Bitcoin mining is vulnerable. *Communications of the ACM*, *61*(7), 95–102. https://doi.org/10.1145/3212998

Furlanello, T., Lipton, Z., Tschannen, M., Itti, L., & Anandkumar, A. (2018). Born again neural networks. *Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research*, *80*, 1607–1616. https://proceedings.mlr.press/v80/furlanello18a.html

Grunspan, C., & Pérez-Marco, R. (2018a). Double spend races. *International Journal of Theoretical and Applied Finance*, *21*(08), 1850053. https://doi.org/10.1142/s021902491850053x

Grunspan, C., & Pérez-Marco, R. (2018b). *On profitability of selfish mining*. arXiv preprint arXiv:1805.08281. https://arxiv.org/abs/1805.08281

Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the knowledge in a neural network*. arXiv preprint arXiv:1503.02531. https://arxiv.org/abs/1503.02531

mempool.space. (2026). *Bitcoin block data, heights 840000 to 844031 [data set]*. https://mempool.space/api/v1/blocks/

Merton, R. K. (1938). Social structure and anomie. *American Sociological Review*, *3*(5), 672. https://doi.org/10.2307/2084686

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. https://bitcoin.org/bitcoin.pdf

Pereira, T., Nascimento, E., Resck, L. E., Mesquita, D., & Souza, A. (2023). Distill n' explain: Explaining graph neural networks using simple surrogates. *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research*, *206*, 6199–6214. https://proceedings.mlr.press/v206/pereira23a.html

Sutherland, E. H. (1947). *Principles of criminology* (4th ed.). Lippincott.

Yu, H., Kaminsky, M., Gibbons, P. B., & Flaxman, A. (2006). SybilGuard: Defending against sybil attacks via social networks. *ACM SIGCOMM Computer Communication Review*, *36*(4), 267–278. https://doi.org/10.1145/1151659.1159945

Author information

Affiliation: Omxus Press. All authors wrote the paper. The authors declare no competing interests. Correspondence: research@omxus.com.