RESEARCH STATEMENT

KENNY GILL

My research is in the interaction between combinatorics and the theory of computation, broadly speaking. Currently I have two ongoing projects: in one I investigate the computational difficulty of Ramsey-theoretic instance-solution problems, and in the other I develop a new complexity measure for finite strings using probabilistic automata and study its properties.

1. Indivisibility and Weihrauch complexity

This project lies in the intersection of Ramsey theory, reverse mathematics, computability, and computable structure theory. A countable structure $\mathcal{S}$ is *indivisible* if for every presentation $\mathcal{A}$ of $\mathcal{S}$ (i.e., every isomorphic copy of $\mathcal{S}$, represented as an infinite binary sequence) and every coloring c of the elements of $\mathcal{A}$ using finitely many colors, there is a monochromatic substructure isomorphic to $\mathcal{S}$. Both $(\mathbb{N}, <)$ and $(\mathbb{Q}, <)$ are indivisible; the indivisibility of $\mathbb{N}$ is just the infinite pigeonhole principle. On the other hand, $(\mathbb{Z}, <)$ is not indivisible because if one colors all positive numbers red and all negative numbers blue, then every monochromatic set has either a greatest or a least element. Many other indivisible structures exist, such as

- $\mathcal{R}$, the Rado graph;
- H_n, the universal countable homogeneous n-clique-free graph (Henson graph), for each n;
- $\mathscr{E}^n$, $n-1$ infinitely refining equivalence relations on $\mathbb{N}^n$;
- Many other Fraïssé limits; ordinals of the form ω^α; nonscattered linear orders; etc.

My overarching goal is to investigate the computational difficulty of finding a substructure $\mathcal{W}$ witnessing the indivisibility of $\mathcal{S}$, given $\mathcal{A}$ and c, and to compare this difficulty among various $\mathcal{S}$. For instance, some structures are *computably indivisible* (CI), meaning there is always some $\mathcal{W}$ computable from $\mathcal{A}$ and c. Then $\mathbb{N}$, $\mathbb{Q}$, $\mathcal{R}$, and ω^α are CI, but I showed that H_n isn't for any $n \geq 3$ [3]; neither is $\mathscr{E}^2$, as follows from another of my results [4, Theorem 5.1].[1] However, it arguably seems like there should be a more nuanced classification even among CI structures. Is it really just as easy to compute a monochromatic subcopy of $\mathbb{Q}$ from a given coloring as it is for $\mathbb{N}$?

Weihrauch reducibility. We can make this vague question precise using Weihrauch reducibility, a framework originating in computable analysis in the late 1980s which has been widely used in reverse mathematics since the early 2010s (although its specific application to indivisibility of structures other than $\mathbb{N}$ is new to my dissertation). A *problem* is a partial multivalued function P on Baire space $\mathbb{N}^\mathbb{N}$, viewed as mapping instances of the problem to solutions of the instance. For example, if $\mathcal{S}$ is an indivisible structure, then its *indivisibility problem* $\mathsf{Ind}\,\mathcal{S}$ is the problem whose instances are presentations $\mathcal{A}$ of $\mathcal{S}$ together with colorings c of $\mathcal{A}$, and for which a solution to $\langle \mathcal{A}, c\rangle$ is any c-monochromatic subcopy $\mathcal{W}$ of $\mathcal{A}$, identified with its characteristic function. We can think of $\mathsf{Ind}\,\mathcal{S}$ as a "black box" capable of finding a suitable $\mathcal{W}$ given any coloring and presentation. One can also consider $\mathsf{Ind}\,\mathcal{S}_k$ which restricts $\mathsf{Ind}\,\mathcal{S}$ to colorings bounded by k. Next, if P and Q are any two problems, then P is *Weihrauch reducible* to Q (written $P \leq_\mathrm{W} Q$) if, intuitively speaking, every instance of P can be solved by a single application of Q, up to some uniformly computable "glue" needed to translate between P and Q. This glue takes the form of Turing functionals Δ and Ψ, where Δ transforms instances of P to instances of Q and Ψ translates solutions of Q to solutions of P (Ψ additionally has oracle access to the original instance of P; see [1] for details).

We have $\mathsf{Ind}\,\mathbb{N} \leq_\mathrm{W} \mathsf{Ind}\,\mathbb{Q}$: if c is any coloring of $\mathbb{N}$, and we simply copy c over to $\mathbb{Q}$ by identifying rationals with their indices in some enumeration of $\mathbb{Q}$, then $\mathsf{Ind}\,\mathbb{Q}$ finds a densely ordered monochromatic

[1] That $\mathscr{E}^2$ is not CI was already shown in unpublished work of Ackerman, Freer, Reimann, and Westrick by a direct construction, but the cited result strengthens this.

set—in particular an infinite monochromatic set, which is enough to solve $\mathsf{Ind}\,\mathbb{N}$. However, it turns out that $\mathsf{Ind}\,\mathbb{Q} \not\leq_{\mathrm{W}} \mathsf{Ind}\,\mathbb{N}$, as follows from independent work of several authors (including myself). Hence it is, in a completely rigorous sense, strictly harder to find a monochromatic subcopy of $\mathbb{Q}$ than it is for $\mathbb{N}$.

My research program. Weihrauch reducibility can reveal subtle distinctions between combinatorial principles which may not be otherwise apparent. My work as represented in [4, 6] mainly focused on teasing apart these distinctions among the indivisibility problems for several particular structures, separating them from some standard benchmark problems in the Weihrauch lattice and from other problems studied in reverse mathematics. For example, I showed [4, Theorem 5.1] that the Weihrauch degree of $\mathsf{Ind}\,\mathscr{E}^2$ is strictly between those of two widely studied versions of Ramsey's theorem for pairs, RT^2 and SRT^2, whose logical separation was a major open problem in reverse mathematics until a few years ago (see e.g. [2] for details). $\mathsf{Ind}\,\mathscr{E}^2$ appears to be the first known combinatorial principle with this property. I also showed that $\mathsf{Ind}\,\mathbb{Q}$ is too weak to solve the benchmark problem $\mathsf{C}_{\mathbb{N}}$, which is equivalent to finding a monochromatic rational interval when given a coloring for which there is one [4, Theorem 4.1]; but on the other hand that it cannot be solved by any problem which is "c.e. guessable" in a suitable sense, including $\mathsf{C}_{\mathbb{N}}$ [4, Theorem 4.7].

There are many further avenues to explore with regard to specific structures, e.g., clarifying the relationship between $\mathbb{Q}$ and $\mathcal{R}$ and between $\mathscr{E}^n$ and ω^n. However, my overriding interest right now is to develop a more general framework with the aim of understanding properties shared by all indivisibility problems, or at least by large classes of them. For instance, every indivisibility problem is a cylinder, meaning that one can always solve $\mathsf{Ind}\,\mathcal{S}$ while simultaneously encoding any given infinite string into every solution. But it quickly becomes difficult to prove very much without additional hypotheses on $\mathcal{S}$—leading me into considerations from computable structure theory, the study of computability-theoretic properties of structures and their isomorphism classes. The indivisibility problems of so-called uniformly computably categorical (u.c.c.) structures, including $\mathbb{Q}$ and $\mathcal{R}$, are particularly well-behaved with respect to $\leq_{\mathrm{W}}$, but there are also many non-u.c.c. structures $\mathcal{S}$ with $\mathsf{Ind}\,\mathcal{S} \equiv_{\mathrm{W}} \mathsf{Ind}\,\mathbb{Q}$, such as $n \cdot \mathbb{Q}$ and any linear order with finitely many adjacent elements. More investigation is needed to find precisely when such rigidity can occur, for $\mathbb{Q}$ and others.

Most recently I have been exploring product operations defined between arbitrary structures in an attempt to generalize constructions found in the literature, with the aim of capturing "the right amount" of information uniformly computable from a presentation of the product—while playing nice with $\leq_{\mathrm{W}}$ and preserving other desirable properties like being indivisible or u.c.c. Also interesting are variations on $\mathsf{Ind}\,\mathcal{S}$ such as only outputting the color of a solution, restricting to a fixed presentation of $\mathcal{S}$, etc.

2. Probabilistic automatic complexity

The Kolmogorov complexity as a string function is well-known to be noncomputable. This has motivated a number of authors to introduce computable string complexity measures using weaker models of computation than a general Turing machine. Particularly appealing for their hands-on combinatorial flavor are those due to Shallit and Wang [9] and Hyde [7], which use DFAs and NFAs, respectively. The DFA complexity $A_D(x)$ is the least number of states of a DFA which uniquely accepts x among strings of length $|x|$. The NFA complexity $A_N(x)$ is analogous but with the extra requirement that the witnessing NFA have a unique accepting path of length $|x|$. (See e.g. [8] for more details.)

In [6, 5], I introduced a new complexity measure in the same spirit but based on probabilistic finite-state automata (PFAs). A PFA is like a DFA where the state transitions have probabilities, and so each word x is given a probability of acceptance $\rho(x)$ rather than a binary acceptance or rejection as with a DFA or NFA. Then instead of asking for x to be the unique string accepted of its length, we ask for it to be the unique string most likely to be accepted. It will be convenient to define the *gap function* of a PFA M as $\mathrm{gap}_M(x) = \min\{\, \rho_M(x) - \rho_M(y) : |y| = |x| \text{ and } y \neq x \,\}$. This measures the degree to which M separates x from other strings of the same length, with a positive gap being equivalent to x having the unique highest probability:

Definition 2.1. The PFA complexity of a string x, $A_P(x)$, is the least number of states of a PFA M with $\mathrm{gap}_M(x) > 0$.

A_D and A_N are computable since there are only finitely many DFAs or NFAs of a given size over a given alphabet, and one can carry out a brute-force search of automata with up to $|x|+1$ states (or less for NFAs). A_P also turns out to be computable, although the reason is not so obvious; a model-theoretic proof was suggested by Bjørn Kjos-Hanssen and appears in [5].

A_P generally seems to be much lower than A_N. For instance, as shown by Hyde, the strings with $A_N = 2$ are exactly those of the form $i^m j$, ij^m, $(ij)^m$, or $(ij)^m i$ for $m \geq 1$, but

Theorem 2.2. *If $x \in \{i,j\}^*$, then $A_P(x) = 2$ if and only if x is of the form*

$$i^n j^m, \qquad i^n j^m i, \qquad i^n (ji)^m, \qquad \text{or} \quad i^n (ji)^m j \qquad \text{for some } n, m \geq 1.$$

(We have $A_P(i^n) = A_N(i^n) = 1$.) The proof of this theorem in [5] actually shows that a single PFA usually witnesses the complexity of an infinite family of strings of similar structure, e.g., $\{\, 0^n 1^m : m \geq 1 \,\}$ for a fixed n. But in general, one heuristically expects $\mathrm{gap}(x)$ to decrease as $|x|$ increases. This can be seen through the lens of the correspondence between PFAs and affine iterated function systems (IFSs) exploited in the proof of Theorem 2.2: all associated IFSs are contractive, so all orbits converge exponentially to the attractor. Then it may be that a witness for $A_P(x)$ does not distinguish x especially well from other strings of its length, in that their acceptance probabilities are very close together. We can try to get around this by introducing a required lower bound on $\mathrm{gap}(x)$:

Definition 2.3. The PFA complexity of x with gap γ, $A_{P,\gamma}(x)$, is the least number of states of a PFA M such that $\mathrm{gap}_M(x) > \gamma$. Here γ may be any element of $[0,1)$. (Hence $A_{P,0} = A_P$.)

Then $A_{P,\gamma}$ is computable for all $\gamma \geq 0$, and in fact uniformly computable almost everywhere as a function of x and γ [5, Section 5]. (How *efficiently* they can be computed is open, but the proof uses real quantifier elimination, so it would be quite surprising to do better than the state of the art there.)

Theorem 2.2 is proven by an exhaustive analysis of the dynamics of IFSs associated with 2-state PFAs which would almost certainly be infeasible to imitate for larger PFAs, or even for 2-state PFAs over a larger alphabet. The original motivation of that theorem was to show that A_P can be greater than 2, but it is still unknown how large it can get:

Question 2.4. Is A_P unbounded? If not, what is its maximum value (overall or over a given alphabet)?

A brute-force search using SageMath has shown that $A_P(x) \leq 3$ for all $x \in \{0,1\}^{\leq 10}$. I suspect that $A_P(x) \leq 3$ for all x, but that $A_{P,\gamma}(x)$ is unbounded for all $\gamma > 0$. The most promising attacks on the second conjecture have attempted to exploit the observation mentioned above that for a generic PFA M witnessing the complexity of each member of a sequence (x_n), one expects $\mathrm{gap}_M(x_n) \to 0$ as $n \to \infty$. But this is the case only generically, not always, and one runs into technical issues when making any such argument precise. Further insights into the behavior of $\mathrm{gap}_M(x)$ and its interaction with the topology of the space of PFAs would be helpful for this (and of independent interest). If A_P or $A_{P,\gamma}$ turn out to be unbounded, one could define a notion of fractal dimension of finite or infinite sequences using either quantity and see if it could be related to existing such notions which use finite automata and transducers. One could study the set of A_P-random strings, i.e., those with highest possible complexity for their length. There are also several other approaches to PFA complexity floated in [5] whose properties are worth investigating.

References

[1] V. Brattka, G. Gherardi, and A. Pauly, Weihrauch complexity in computable analysis. In *Handbook of computability and complexity in analysis* (V. Brattka and P. Hertling, eds.), Springer, 2021, pp. 367–417. arXiv: 1707.03202.

[2] D. Dzhafarov and C. Mummert, *Reverse mathematics: Problems, reductions, and proofs.* Springer, 2022.

[3] K. Gill, A note on the indivisibility of the Henson graphs, submitted (2024). arXiv: 2310.20097.

[4] K. Gill, Indivisibility and uniform computational strength, submitted (2024). arXiv: 2312.03919.

[5] K. Gill, Probabilistic automatic complexity of finite strings, submitted (2024). arXiv: 2402.13376.

[6] K. Gill, *Two studies in complexity.* Ph.D. dissertation, The Pennsylvania State University, 2023.

[7] K. Hyde, *Nondeterministic finite state complexity*, MA thesis, University of Hawai'i, Manoa, 2013.

[8] B. Kjos-Hanssen, *Automatic complexity: A computable measure of irregularity.* De Gruyter, 2024.

[9] J. Shallit and M.-w. Wang, Automatic complexity of strings, *J. Autom. Lang. Comb.* **6**(4) (2001), 537–554.