Two Departures from Classical Information Theory

Blake C. Stacey
QBism Group, Physics Department
University of Massachusetts Boston

26 June 2018
Happy to be Back!

Multiscale Structure in Eco-Evolutionary Dynamics

Blake C. Stacey
The Story of My Life

What happens when there are patterns among *three* things that aren't visible from the relations between *two* of them?
Classical Complex Interdependencies

\[ a \]

\[ b \]

\[ c \]

\[ 0 \]

\[ 1 \]

\[ -1 \]

\[ 1 \]

\[ 0 \]

\[ 1 \]

\[ 0 \]

\[ arXiv:1705.03927 \]
How Quantum is Quantum?

In complex systems, there can be “more going on inside” a random variable than we’re used to.

But what if there is less?
We want $d^2$ vectors in a $d$-dimensional complex vector space such that

$$\left|\langle \pi_j | \pi_k \rangle \right|^2 = \frac{d \delta_{jk} + 1}{d + 1}.$$ 

In terms of the projectors $\Pi_j = |\pi_j \rangle \langle \pi_k|$, 

$$\text{tr} \, \Pi_j \Pi_k = \frac{d \delta_{jk} + 1}{d + 1}.$$
We want $d^2$ vectors in a $d$-dimensional complex vector space such that

$$|\langle \pi_j | \pi_k \rangle|^2 = \frac{d\delta_{jk} + 1}{d + 1}.$$ 

In terms of the projectors $\Pi_j = |\pi_j\rangle\langle \pi_k|$, 

$$\text{tr} \Pi_j \Pi_k = \frac{d\delta_{jk} + 1}{d + 1}.$$ 

Solutions are known up through $d = 151$ and assorted higher $d$, but in general the problem remains open.
A SIC in $d = 2$
Table 1. References classified by the topics to which they give significant coverage.

<table>
<thead>
<tr>
<th>Topic</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract algebra</td>
<td>[8,10,11,70]</td>
</tr>
<tr>
<td>Algebraic number theory</td>
<td>[5–7,26,61,71]</td>
</tr>
<tr>
<td>Category theory</td>
<td>[72,73]</td>
</tr>
<tr>
<td>Compressed sensing and signal processing</td>
<td>[74–77]</td>
</tr>
<tr>
<td>Elliptic curves</td>
<td>[40,49,78,79]</td>
</tr>
<tr>
<td>Exact solutions</td>
<td>[4,8,25,26,80–85]</td>
</tr>
<tr>
<td>Frame theory</td>
<td>[25,36,62,76,77,86–101]</td>
</tr>
<tr>
<td>Finite group theory</td>
<td>[9,12,13,21,46,100,102–107]</td>
</tr>
<tr>
<td>Generalized and approximate SICs</td>
<td>[70,73,108–116]</td>
</tr>
<tr>
<td>Historical overview</td>
<td>[44,117,118]</td>
</tr>
<tr>
<td>Informational power and entropy</td>
<td>[13,21,52,119–129]</td>
</tr>
<tr>
<td>Laboratory experiments</td>
<td>[27–34]</td>
</tr>
<tr>
<td>Multipartite systems and sequential measurements</td>
<td>[46,82,130–136]</td>
</tr>
<tr>
<td>Quantum communication and cryptography</td>
<td>[27,28,137–143]</td>
</tr>
<tr>
<td>Quantum computing and contextuality</td>
<td>[13,20,39,52,135,144–151]</td>
</tr>
<tr>
<td>Quantum decoherence</td>
<td>[152,153]</td>
</tr>
<tr>
<td>Quantum entanglement</td>
<td>[19,112–114,137,139,154]</td>
</tr>
<tr>
<td>Quantum tomography</td>
<td>[18,30,35,62,70,133,134,138,140,155–165]</td>
</tr>
<tr>
<td>Quaternions and octonions</td>
<td>[9,13,47,48,72,166,167]</td>
</tr>
<tr>
<td>Reconstructing quantum theory</td>
<td>[14,17,22,23,39,73,117,135,168–172]</td>
</tr>
<tr>
<td>SICs and Mutually Unbiased Bases</td>
<td>[16,52,79,125,145,151,173–186]</td>
</tr>
<tr>
<td>SICs, Wigner functions and phase space representations</td>
<td>[20,28,39,52,82,88,89,144,174,187–190]</td>
</tr>
</tbody>
</table>
In general:

- Tomographically optimal
- Maximally sensitive to eavesdropping
In general:

- Tomographically optimal
- Maximally sensitive to eavesdropping
- Best possible *reference measurements* for probabilistic representation of quantum theory
In general:
- Tomographically optimal
- Maximally sensitive to eavesdropping
- Best possible *reference measurements* for probabilistic representation of quantum theory

In specific cases:
- Nice entanglement properties ($d = 4$)
SICs — Nice Properties

In general:

- Tomographically optimal
- Maximally sensitive to eavesdropping
- Best possible reference measurements for probabilistic representation of quantum theory

In specific cases:

- Nice entanglement properties ($d = 4$)
- Algebraic number theory ($d \geq 4$)
SICs — Nice Properties

In general:
- Tomographically optimal
- Maximally sensitive to eavesdropping
- Best possible *reference measurements* for probabilistic representation of quantum theory

In specific cases:
- Nice entanglement properties \((d = 4)\)
- Algebraic number theory \((d \geq 4)\)
- Finite group theory \((d = 2, 3, 8)\)
SICs — Nice Properties

In general:

▶ Tomographically optimal
▶ Maximally sensitive to eavesdropping
▶ Best possible reference measurements for probabilistic representation of quantum theory

In specific cases:

▶ Nice entanglement properties \((d = 4)\)
▶ Algebraic number theory \((d \geq 4)\)
▶ Finite group theory \((d = 2, 3, 8)\)
▶ Maximally magic resources for QC \((d = 2, 3, 8)\)
Take $\omega = e^{2\pi i/3}$, and construct the set of states $\{|\pi_j\rangle\}$ given by the columns of

$$
\frac{1}{\sqrt{2}} \begin{pmatrix}
0 & -1 & 1 & 0 & -1 & 1 & 0 & -1 & 1 \\
1 & 0 & -1 & \omega & 0 & -\omega & \omega^2 & 0 & -\omega^2 \\
-1 & 1 & 0 & -\omega^2 & \omega^2 & 0 & -\omega & \omega & 0
\end{pmatrix}.
$$

Easy to recover by remembering the first column and applying

$$
X|n\rangle = |n + 1\rangle, \quad Z|n\rangle = \omega^n|n\rangle.
$$
$d = 3$: Hesse SIC

Nice properties:
- Not just group covariant, but doubly transitive
- ... which makes quantities like the triple products $\text{tr} \prod_j \prod_k \prod_l$ simplify drastically!

The stabilizer subgroup of the fiducial vector goes by many names:
- binary tetrahedral group $Q \rtimes \mathbb{Z}_3$
- $SL(2, 3)$
- ... 24 units of the Hurwitz integers
Quantum Probability

In quantum information, a measurement is a POVM:

$$\{E_i\}, \text{ such that } \sum_i E_i = I.$$  

The probability of obtaining outcome $i$ is

$$p(E_i) = \text{tr}(\rho E_i).$$

A SIC forms a particularly special POVM $\{H_i\}$ by scaling:

$$H_i = \frac{1}{d} \Pi_i.$$  

Given the $p(H_i)$, we can reconstruct $\rho$:

$$\rho = \sum_i \left[ (d + 1)p(H_i) - \frac{1}{d} \right] \Pi_i.$$
A Picture for the Math

\[ P(H_i) \]

\[ \{H_i\} \]

\[ \sigma_i \]

\[ P(D_j | H_i) \]

\[ \{D_j\} \]

\[ Q(D_j) \]

\[ \rho \]
Quantum Probability

Classical propagation of probabilities:

\[ p(D_j) = \sum_{i} p(H_i) p(D_j|H_i). \]
Quantum Probability

Classical propagation of probabilities:

\[ p(D_j) = \sum_i p(H_i) p(D_j|H_i). \]

But because

\[ \rho = \sum_i \left[ (d + 1)p(H_i) - \frac{1}{d} \right] \Pi_i , \]

the quantum answer is

\[ p(D_j) = \text{tr}(\rho D_j) = \sum_i \left[ (d + 1)p(H_i) - \frac{1}{d} \right] p(D_j|H_i). \]
SICs are the Best ICs
SICs are the Best ICs

In general, we’ll have

\[ p(D) = p(D|H)\Phi p(H), \]

where \( \Phi = I \) would be the classical relation.

**Theorem**

*Given any Minimal IC,*

\[ \| I - \Phi \| \geq \| I - \Phi_{\text{SIC}} \| \]

*for any unitarily invariant norm \( \| \cdot \| \), with equality iff the MIC is a SIC.*

arXiv:1805.08721
Quantum Probability

A state $\rho$ is pure iff

$$\text{tr}\rho = \text{tr}\rho^2 = \text{tr}\rho^3 = 1.$$  

When we use a SIC to turn $\rho$ into $p$, the quadratic condition becomes

$$\sum_{i=1}^{d^2} p(i)^2 = \frac{2}{d(d + 1)}.$$
Quantum Probability

A state $\rho$ is pure iff

$$\text{tr}\rho = \text{tr}\rho^2 = \text{tr}\rho^3 = 1.$$  

When we use a SIC to turn $\rho$ into $p$, the quadratic condition becomes

$$\sum_{i=1}^{d^2} p(i)^2 = \frac{2}{d(d + 1)}.$$  

Nifty for me, because

$$N_{\text{eff}} = \binom{d + 1}{2}.$$
Quantum Probability

And the $\text{tr} \rho = 1$ condition? Well, ...
And the $\text{tr} \, \rho = 1$ condition? Well, ... 

$$
\sum_{ijk} C_{ijk} p(i)p(j)p(k) = \frac{d + 7}{(d + 1)^3},
$$

where

$$
C_{ijk} = \text{Re} \text{tr} \Pi_j \Pi_k \Pi_l.
$$
And the $\text{tr} \rho = 1$ condition? Well, . . .

$$\sum_{ijk} C_{ijk} p(i)p(j)p(k) = \frac{d + 7}{(d + 1)^3},$$

where

$$C_{ijk} = \text{Re} \text{tr} \Pi_i \Pi_j \Pi_k.$$

But! Using the Hesse SIC,

$$\sum_i p(i)^3 - 3 \sum_{(ijk) \in S} p(i)p(j)p(k) = 0.$$
Quantum Probability

Using the Hesse SIC,

$$\sum_{i} p(i)^3 - 3 \sum_{(ijk) \in S} p(i)p(j)p(k) = 0.$$ 

The set $S$ is a *steiner triple system*:

$$\begin{align*}
1 & \quad 2 & \quad 3 \\
4 & \quad 5 & \quad 6 ; & \quad \text{reading off the lines,} & \quad S = \\
7 & \quad 8 & \quad 9
\end{align*}$$

$$\begin{pmatrix}
(123) & (456) & (789) \\
(147) & (258) & (369) \\
(159) & (267) & (348) \\
(168) & (249) & (357)
\end{pmatrix}.$$
Quantum Probability

A useful fact:

\[
\binom{d}{2} + \binom{d + 1}{2} = d^2.
\]

Theorem

No more than \(d(d - 1)/2\) entries of a quantum state \(p\) can equal zero.

With the Hesse SIC, exactly 12 pure states attain this bound. Each of these 12 states is orthogonal to exactly 3 SIC vectors:

\[
(123) \quad (456) \quad (789) \\
(147) \quad (258) \quad (369) \\
(159) \quad (267) \quad (348) \\
(168) \quad (249) \quad (357)
\]
Quantum Probability

With the Hesse SIC, \textit{exactly} 12 pure states have $d = 3$ zeros. Each of these 12 states is orthogonal to exactly 3 SIC vectors:

\[
\begin{align*}
(123) & \quad (456) & \quad (789) \\
(147) & \quad (258) & \quad (369) \\
(159) & \quad (267) & \quad (348) \\
(168) & \quad (249) & \quad (357)
\end{align*}
\]

These states minimize the Shannon entropy!
Quantum Probability

With the Hesse SIC, *exactly* 12 pure states have $d = 3$ zeros. Each of these 12 states is orthogonal to exactly 3 SIC vectors:

\[
\begin{align*}
(123) & \quad (456) & \quad (789) \\
(147) & \quad (258) & \quad (369) \\
(159) & \quad (267) & \quad (348) \\
(168) & \quad (249) & \quad (357)
\end{align*}
\]

These states minimize the Shannon entropy! Each row is an orthonormal basis, and any two rows are *mutually unbiased.*
With the Hesse SIC, \textit{exactly 12} pure states have $d = 3$ zeros. Each of these 12 states is orthogonal to exactly 3 SIC vectors:

\[
(123) \quad (456) \quad (789) \\
(147) \quad (258) \quad (369) \\
(159) \quad (267) \quad (348) \\
(168) \quad (249) \quad (357)
\]

Each row is an orthonormal basis, and any two rows are \textit{mutually unbiased}.

A complete set of MUB is just what we need to define a \textit{discrete Wigner function} and prove that the Hesse SIC states are \textit{maximally magic}. 
Up to $d = 8$

A Hoggar fiducial:

$$|\pi_0^+\rangle \propto (-1 + 2i, 1, 1, 1, 1, 1, 1, 1, 1)^T.$$  

**Important:** The group this time is the tensor product of three Pauli groups.  
Just for fun — The fiducial stabilizer again has many names:  
- $U_3(3)$ or $PSU(3, 3)$ or $PSU(3, \mathbb{F}_9)$  
- \ldots Or, $G_2(2)^\prime$.  

$G_2(2)$ is the automorphism group of the octavians, or Cayley integers in $\mathbb{O}$.  

Entropy and the Hoggar SIC

A Hoggar fiducial:

$$|\pi_0^+\rangle \propto (-1 + 2i, 1, 1, 1, 1, 1, 1, 1)^T.$$

We can play the same game, finding the states with $d(d - 1)/2 = 28$ zeros. These will again be the Shannon entropy minimizers.
A Hoggar fiducial:

\[ |\pi_0^+\rangle \propto (-1 + 2i, 1, 1, 1, 1, 1, 1, 1)^T. \]

We can play the same game, finding the states with \( d(d-1)/2 = 28 \) zeros. These will again be the Shannon entropy minimizers. This time, they form another SIC!
Entropy and the Hoggar SIC

A Hoggar fiducial:

$$|\pi_0^+\rangle \propto (-1 + 2i, 1, 1, 1, 1, 1, 1, 1, 1)^T.$$  

We can play the same game, finding the states with 
$$d(d - 1)/2 = 28$$ zeros. These will again be the Shannon entropy minimizers. 
This time, they form another SIC! 
It’s another SIC with the same symmetry group. 

$$|\pi_0^-\rangle \propto (-1 - 2i, 1, 1, 1, 1, 1, 1, 1, 1)^T.$$
Entropy and the Hoggar SIC

Each vector in the plus SIC is orthogonal to exactly 28 in the minus SIC, and vice versa.

This leads us into all sorts of fun group theory, and a hidden connection with real equiangular lines!

But today, I want to talk Bell inequalities.
Recall Mermin’s three-qubit Bell inequality. Define

\[ B(\rho) = \langle XXX \rangle - \langle XYY \rangle - \langle YXY \rangle - \langle YYY \rangle. \]

The hypothesis of *local hidden variables* implies

\[-2 \leq B(\rho) \leq 2.\]

Why?
Recall Mermin’s three-qubit Bell inequality. Define

\[ B(\rho) = \langle XXX \rangle - \langle XYY \rangle - \langle YXY \rangle - \langle YYX \rangle. \]

The hypothesis of local hidden variables implies

\[ -2 \leq B(\rho) \leq 2. \]

Why? Imagine the system had hidden properties \((\lambda_1X, \lambda_1Y, \lambda_2X, \lambda_2Y, \lambda_3X, \lambda_3Y)\). Then

\[
\langle XXX \rangle - \langle XYY \rangle - \langle YXY \rangle - \langle YYX \rangle \\
= \lambda_1X_1\lambda_2X_2\lambda_3X_3 - \lambda_1X_1\lambda_2Y_2\lambda_3Y_3 - \lambda_1Y_1\lambda_2X_2\lambda_3Y_3 - \lambda_1Y_1\lambda_2Y_2\lambda_3X_3.
\]
Recall Mermin’s three-qubit Bell inequality. Define

\[ B(\rho) = \langle XXX \rangle - \langle XYY \rangle - \langle YXY \rangle - \langle YYX \rangle. \]

The hypothesis of local hidden variables implies

\[-2 \leq B(\rho) \leq 2.\]

But the GHZ state \( \rho_{\text{GHZ}} \) is an eigenstate of the operator \( XXX \) with eigenvalue \(-1\), and it is also an eigenstate of \( XYY \), of \( YXY \) and of \( YYX \) with eigenvalue \(+1\). Therefore,

\[ |B(\rho_{\text{GHZ}})| = 4. \]
Let’s take our old friend

\[ p(D_j) = \sum_i \left[ (d + 1)p(H_i) - \frac{1}{d} \right] p(D_j | H_i) \]

and write the operator expectation version.

Define

\[ \langle A : i \rangle = \text{tr}(A \Pi_i) \].

Then

\[ \langle A \rangle = (d + 1) \sum_i p(i) \langle A : i \rangle - \frac{1}{d} \sum_i \langle A : i \rangle. \]
For traceless operators (like the Paulis $XYY$, $XYX$, etc.),

$$\langle A \rangle = (d + 1) \sum_i p(i)\langle A : i \rangle.$$ 

We just scale up the “classical answer” by $d + 1$.

Now we get sneaky: The operators in Mermin’s inequality all belong to the group that generates the SIC. Thus,

$$|\langle \psi_i | D_k |\psi_i \rangle|^2 = \frac{1}{d + 1} = \frac{1}{9}.$$
For traceless operators (like the Paulis $XYY$, $XYX$, etc.),

$$\langle A \rangle = (d + 1) \sum_i p(i) \langle A : i \rangle .$$

We just scale up the “classical answer” by $d + 1$.

Now we get sneaky: The operators in Mermin’s inequality all belong to the group that generates the SIC. Thus,

$$|\langle \psi_i \mid D_k \mid \psi_i \rangle|^2 = \frac{1}{d + 1} = \frac{1}{9} .$$

And because each $D_k$ is Hermitian,

$$\langle D_k : i \rangle = \langle \psi_i \mid D_k \mid \psi_i \rangle = \pm \frac{1}{3} .$$
We have shown

\[ \langle D_k : i \rangle = \langle \psi_i | D_k | \psi_i \rangle = \pm \frac{1}{3}. \]

So, the magnitude of Mermin’s \( B \) can never exceed \( 4/3 \). In fact,

\[ |\langle XXX : i \rangle - \langle XYY : i \rangle - \langle YXY : i \rangle - \langle YYX : i \rangle| = \frac{2}{3} \quad \forall \ i. \]
We have shown

\[ \langle D_k : i \rangle = \langle \psi_i | D_k | \psi_i \rangle = \pm \frac{1}{3}. \]

So, the magnitude of Mermin’s $B$ can never exceed $4/3$. In fact,

\[ |\langle XXX : i \rangle - \langle XYY : i \rangle - \langle YXY : i \rangle - \langle YXX : i \rangle| = \frac{2}{3} \quad \forall \ i. \]

And a classical weighted average certainly can’t do better!

… But the GHZ state corresponds to just such an average!
The GHZ state itself corresponds to some probability distribution $p(i)$. Let $\mathcal{O}$ range over the four operators in Mermin’s Bell inequality:

$$\mathcal{O} \in \{XXX, -XYY, -YXY, -YYX\}.$$

For the GHZ state, we have

$$\sum_{\mathcal{O}} \left[ \sum_i p(i) \langle \mathcal{O} : i \rangle \right] = \frac{4}{9}. \quad (1)$$

It is the factor of $(d + 1)$ that lifts us over the edge into nonclassical territory.
References