The impact of the number of averaged attacker's strategies on the results quality in Mixed-UCT

Jan Karwowski[1] and Jacek Mańdziuk[1,2]

[1]Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland
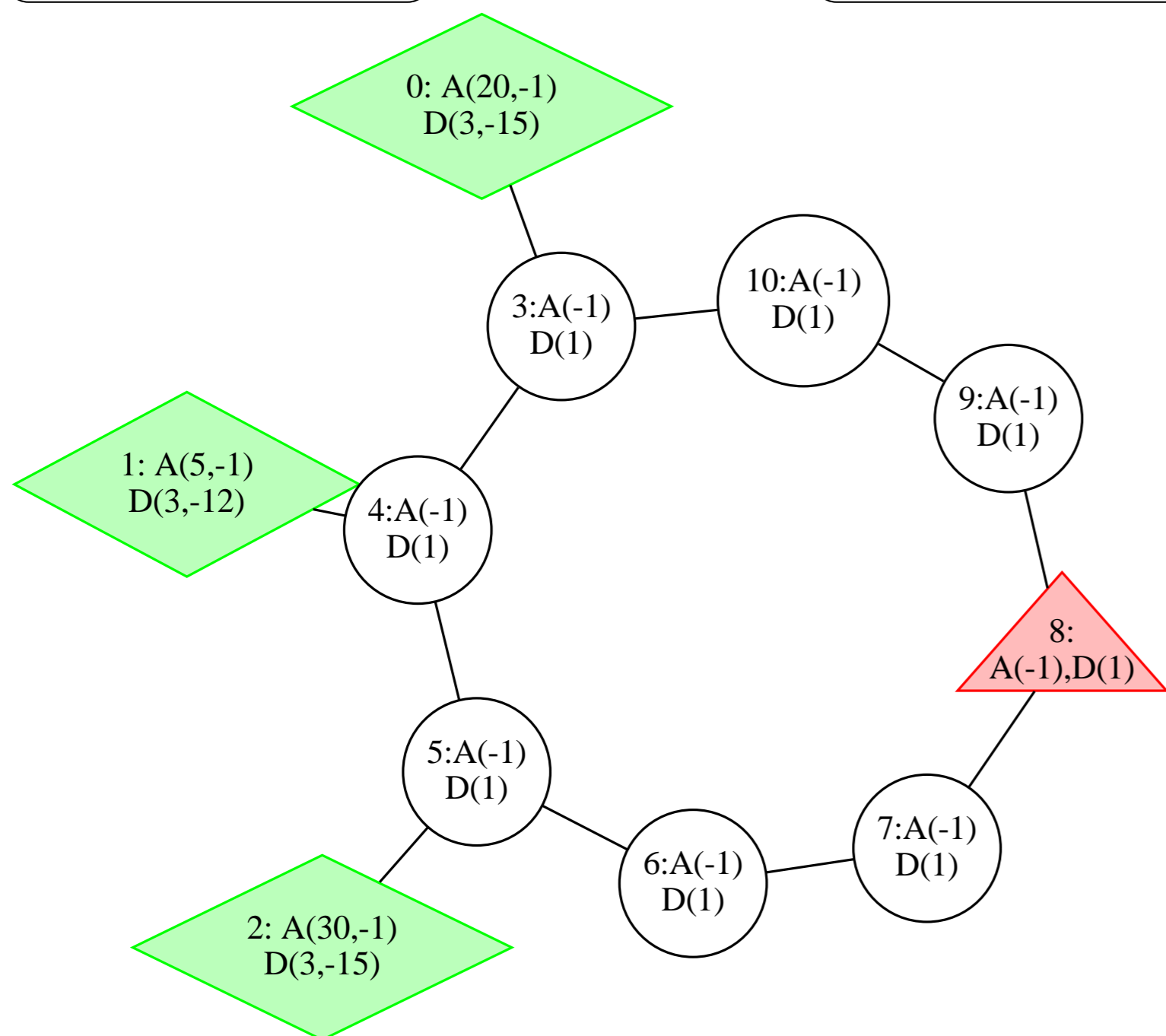{jan.karwowski,j.mandziuk}@mini.pw.edu.pl
[2]School of Computer Science and Engineering, Nanyang Technological University, Singapore

Games on graphs

Security-games inspired

Pursuit-evasion
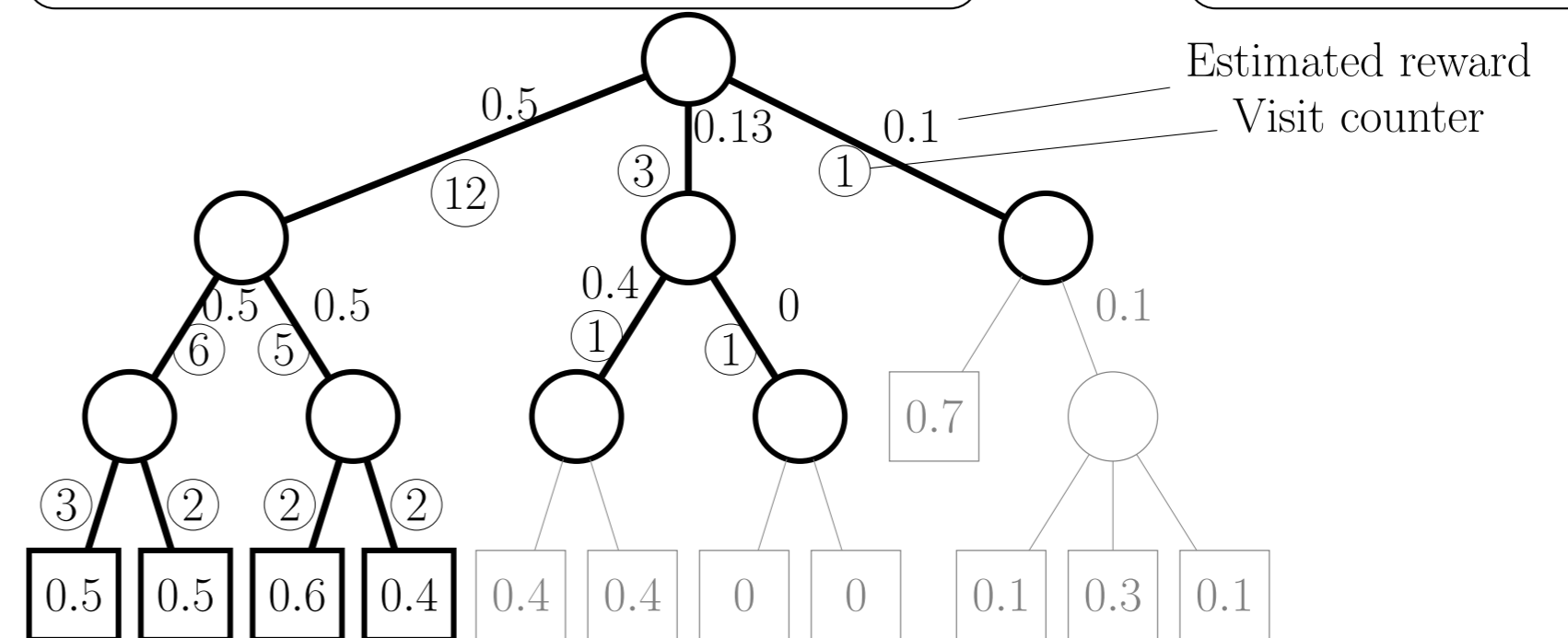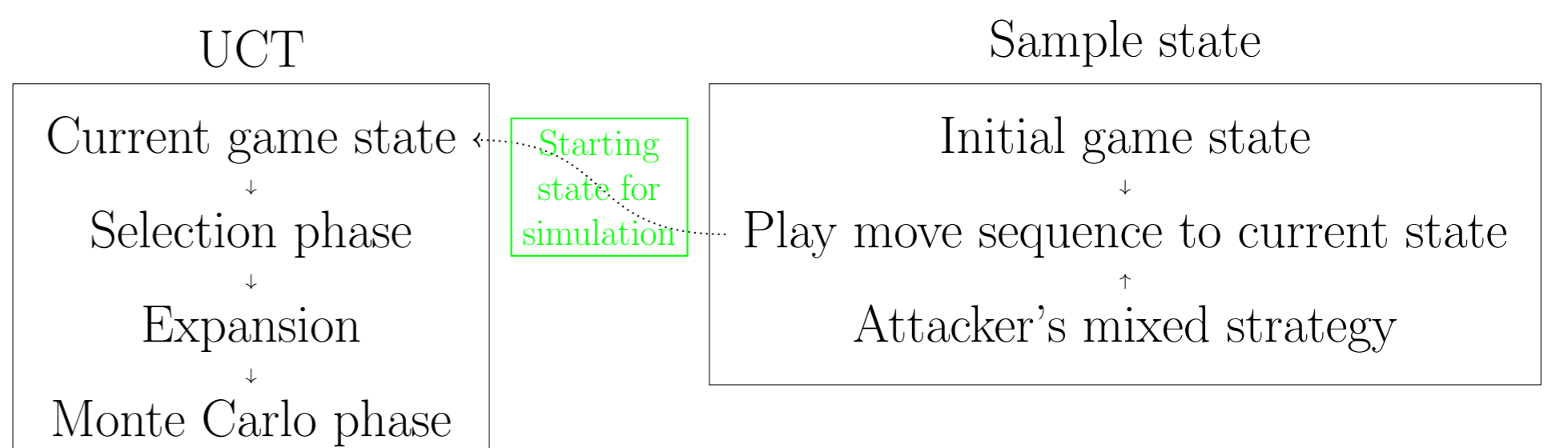
Multi-act games

Imperfect information

Stackelberg Equilibrium

Upper Confidence Bounds Applied to Trees

Uses game rules (compact representation)

Monte-Carlo Tree Search
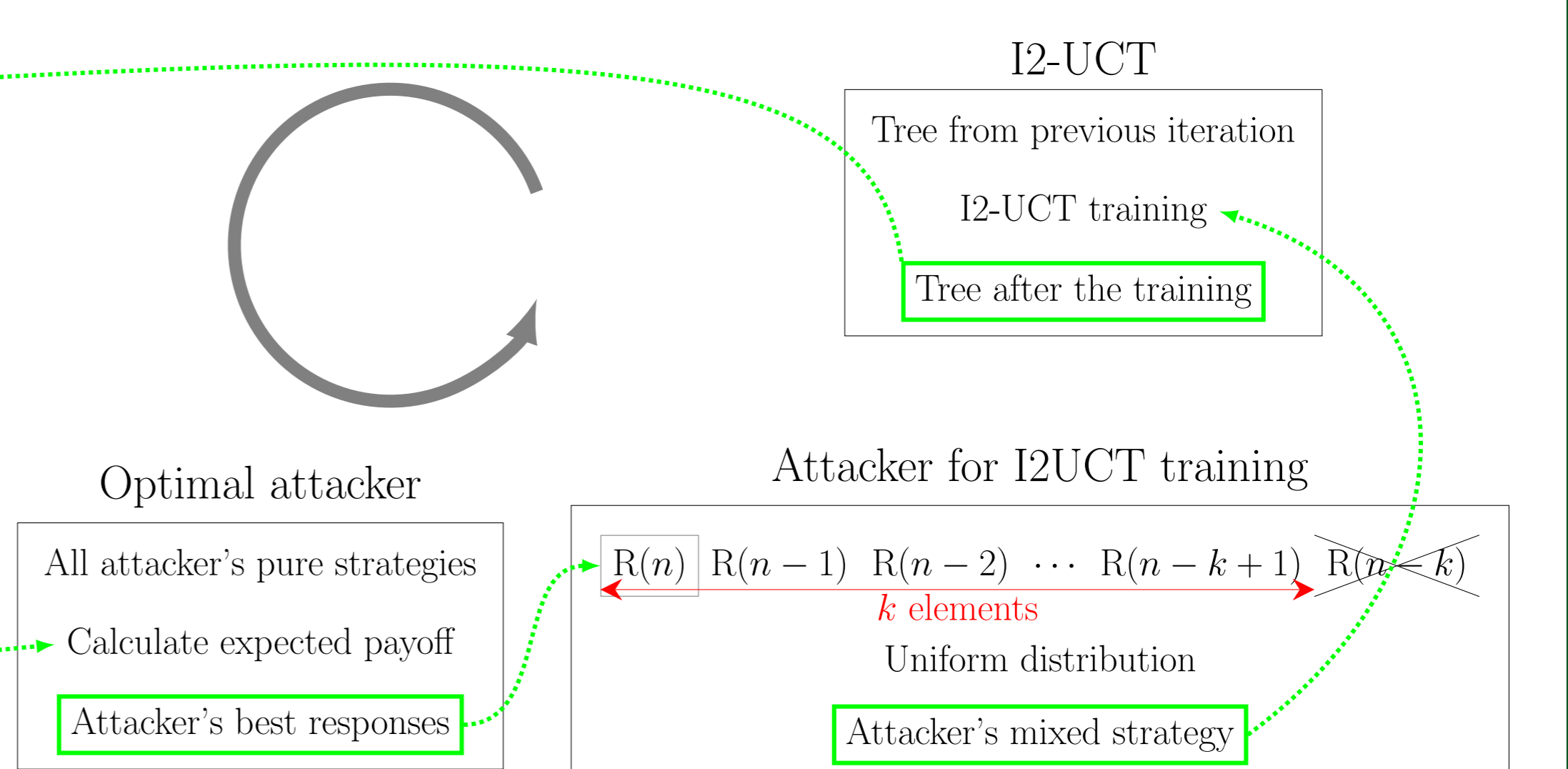
Estimated reward
Visit counter

I2-UCT

UCT	Sample state
Current game state	Initial game state
↓ Selection phase	↓ Play move sequence to current state
↓ Expansion	↑ Attacker's mixed strategy
↓ Monte Carlo phase	

Starting state for simulation

Mixed-UCT

Defender strategy extraction

$$\frac{15}{100} \cdot \frac{5}{15} \cdot \frac{2}{5} =$$

All paths from root to terminal state — Mixed strategy

$p = 0.02$

Optimal attacker
- All attacker's pure strategies
- Calculate expected payoff
- Attacker's best responses

Attacker for I2UCT training

R(n) R(n-1) R(n-2) ⋯ R(n-k+1) R(n-k)
k elements
Uniform distribution
Attacker's mixed strategy

I2-UCT
Tree from previous iteration
↓ I2-UCT training
Tree after the training

Experimental results

game	k = 1			k = 10			k = 100			k = 1000			MILP		U
	R	t	sc	R	t	sc	R	t	sc	R	t	sc	R	t	R
1	-1.57	1448	81%	0.34	1436	98%	0.48	1492	99%	0.47	1611	99%	0.54	28052	-10.46
2	0.03	1049	99%	0.05	1388	100%	0.06	1113	100%	0.07	1351	100%	0.08	106	-7.21
3	-4.32	2319	99%	-4.34	1618	99%	-4.28	1786	100%	-4.28	1135	100%	-4.27	42946	-13.97
3a	-4.5	2003	100%	-4.53	1452	100%	-4.51	1230	100%	-4.5	1487	100%	-4.5	31926	-13.97
3b	2.44	1838	96%	2.57	2020	100%	2.57	1382	100%	2.58	1639	100%	2.58	227	-0.87
3c	-1.52	1777	94%	-1.18	1724	99%	-1.07	1358	100%	-1.06	1169	100%	-1.06	41624	-9.29
3d	0.84	2129	99%	0.69	1595	96%	0.9	1467	100%	0.9	1218	100%	0.9	37870	-4.61
4	-4.87	1777	100%	-4.87	1378	100%	-4.87	1995	100%	-4.87	1310	100%	-4.87	5312	-13.91
4a	-6	1681	100%	-6	1638	100%	-6	1637	100%	-6	1114	100%	-6	5949	-13.84
4b	0.79	1788	100%	0.79	1266	100%	0.79	1639	100%	0.79	1061	100%	0.79	5546	-0.81
4c	-2.85	1783	100%	-2.85	1498	100%	-2.85	1494	100%	-2.85	1203	100%	-2.85	5928	-9.23
4d	0.17	1948	100%	0.17	1455	100%	0.17	1391	100%	0.16	1165	100%	0.17	5155	-4.55

Conclusions

- Increasing history length improves quality of results
- Increasing history length has little impact on computation time

References

[1] Levente Kocsis and Csaba Szepesvári. "Bandit based monte-carlo planning". In: *Machine Learning: ECML 2006*. Springer, 2006, pp. 282–293.

[2] Jan Karwowski and Jacek Mańdziuk. "Mixed Strategy Extraction from UCT Tree in Security Games". In: *ECAI 2016*. IOS Press, 2016, pp. 1746–1747. ISBN: 978-1-61499-672-9.