Playing Hanabi Near-Optimally

Bruno Bouzy

Paris Descartes University

ACG 2017

Leiden – July 3-5, 2017

Outline

- The game of Hanabi
- Previous work
- The hat principle
- Artificial players
- Experiments
- Results
- Conclusions and future work

Hanabi Game Set

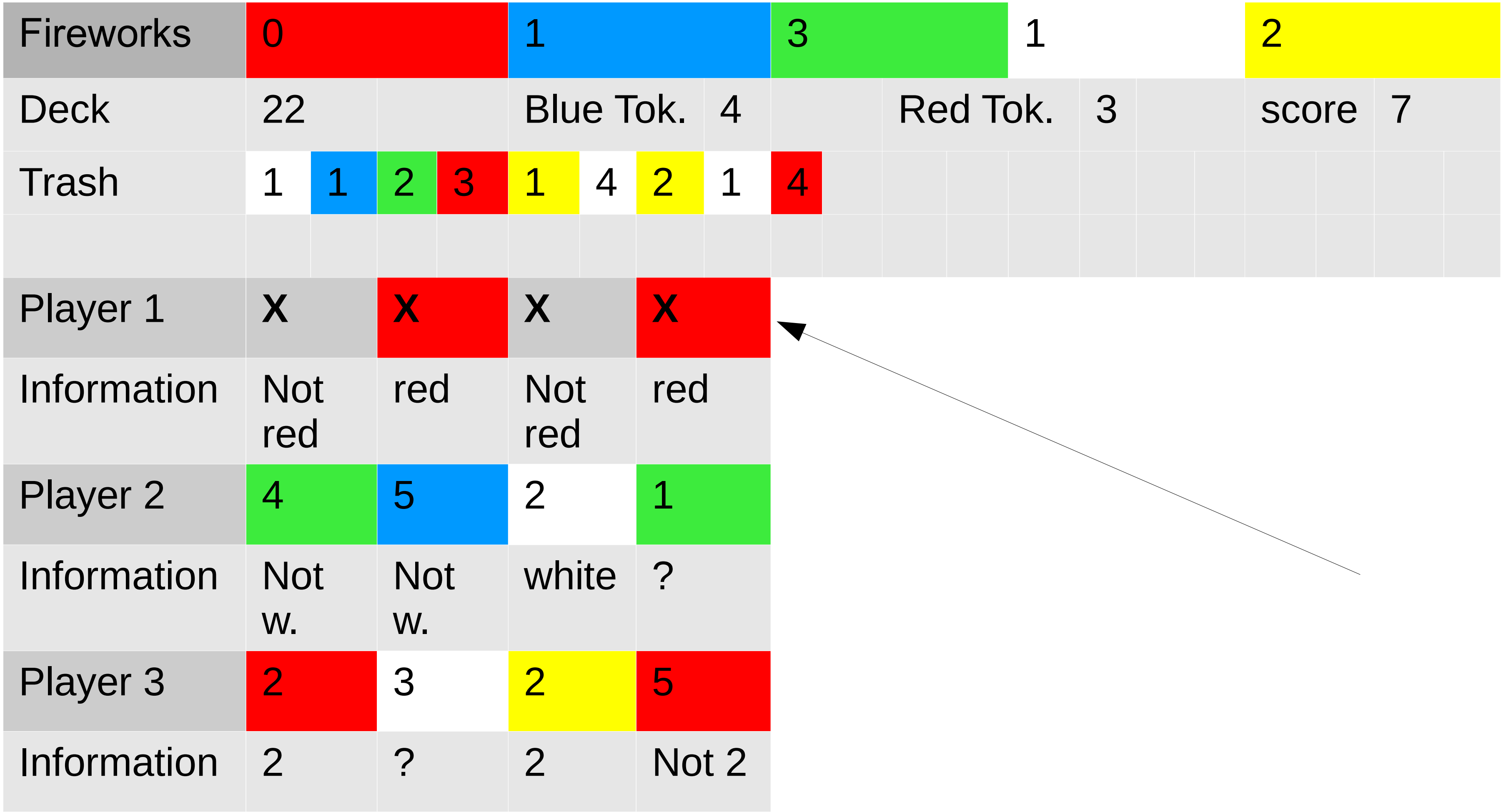
Example

NP=3 players, NCPP=4 cards per player

Fireworks	🟥 0	🟦 1	🟩 3	⬜ 1	🟨 2				
Deck	22	Blue Tok.	4	Red Tok.	3	score	7		
Trash	⬜ 1	🟦 1	🟩 2	🟥 3	🟨 1	⬜ 4	🟨 2	⬜ 1	🟥 4
Player 1	🟦 1	🟥 2	🟩 3	🟥 1					
Information	Not red	red	Not red	red					
Player 2	🟩 4	🟦 5	⬜ 2	🟩 1					
Information	Not w.	Not w.	white	?					
Player 3	🟥 2	⬜ 3	🟨 2	🟥 5					
Information	2	?	2	Not 2					

My own cards are hidden

NP=3 players, NCPP=4 cards per player



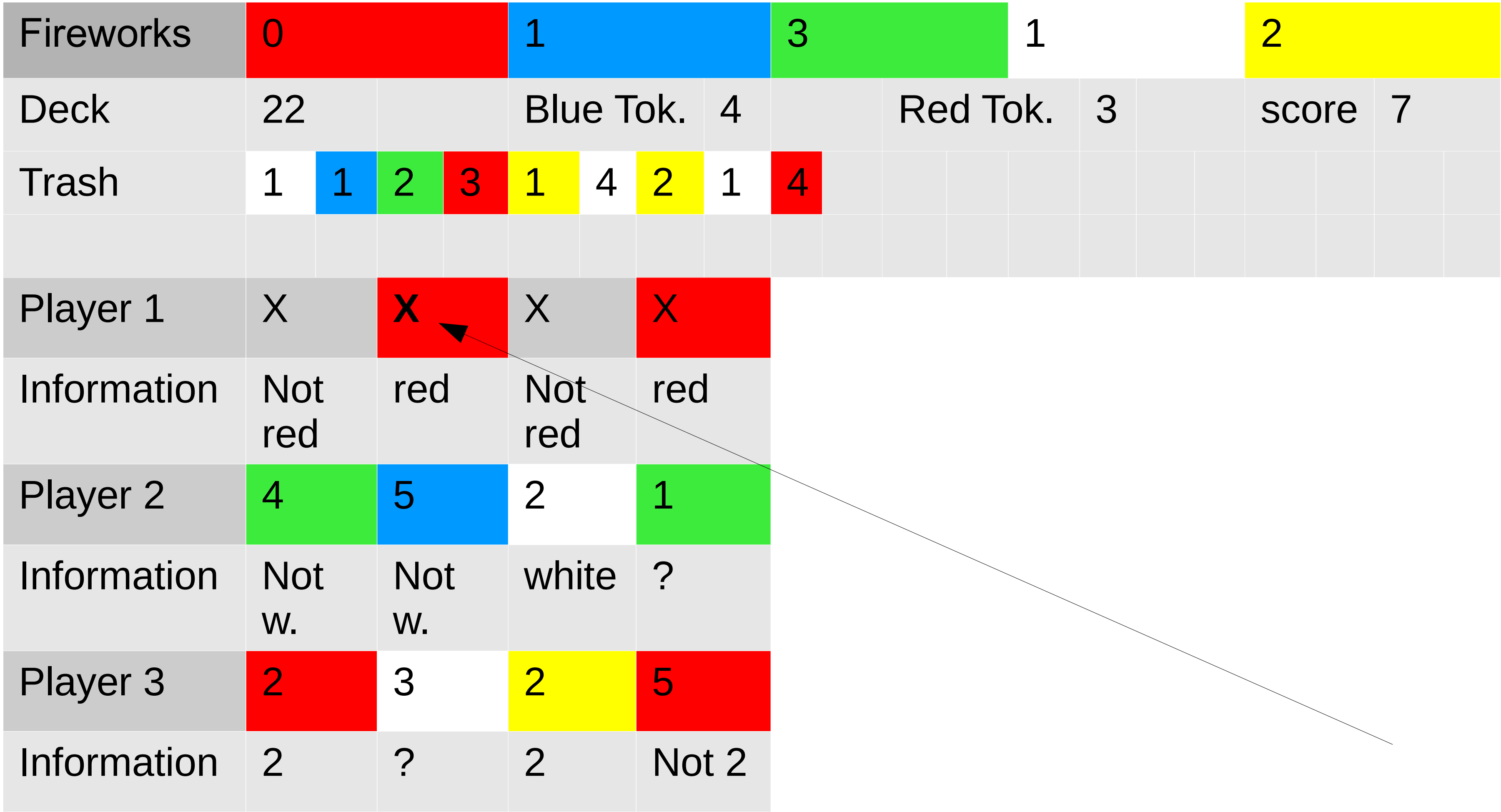
Fireworks	🟥 0	🟦 1	🟩 3	⬜ 1	🟨 2					
Deck	22		Blue Tok.	4		Red Tok.	3		score	7
Trash	⬜ 1	🟦 1	🟩 2	🟥 3	🟨 1	⬜ 4	🟨 2	⬜ 1	🟥 4	
Player 1	X	🟥 X	X	🟥 X						
Information	Not red	red	Not red	red						
Player 2	🟩 4	🟦 5	⬜ 2	🟩 1						
Information	Not w.	Not w.	white	?						
Player 3	🟥 2	⬜ 3	🟨 2	🟥 5						
Information	2	?	2	Not 2						

3 kinds of move

- Play a card
- Discard a card
- Inform a player with either a color or a height

I choose to play card number 2

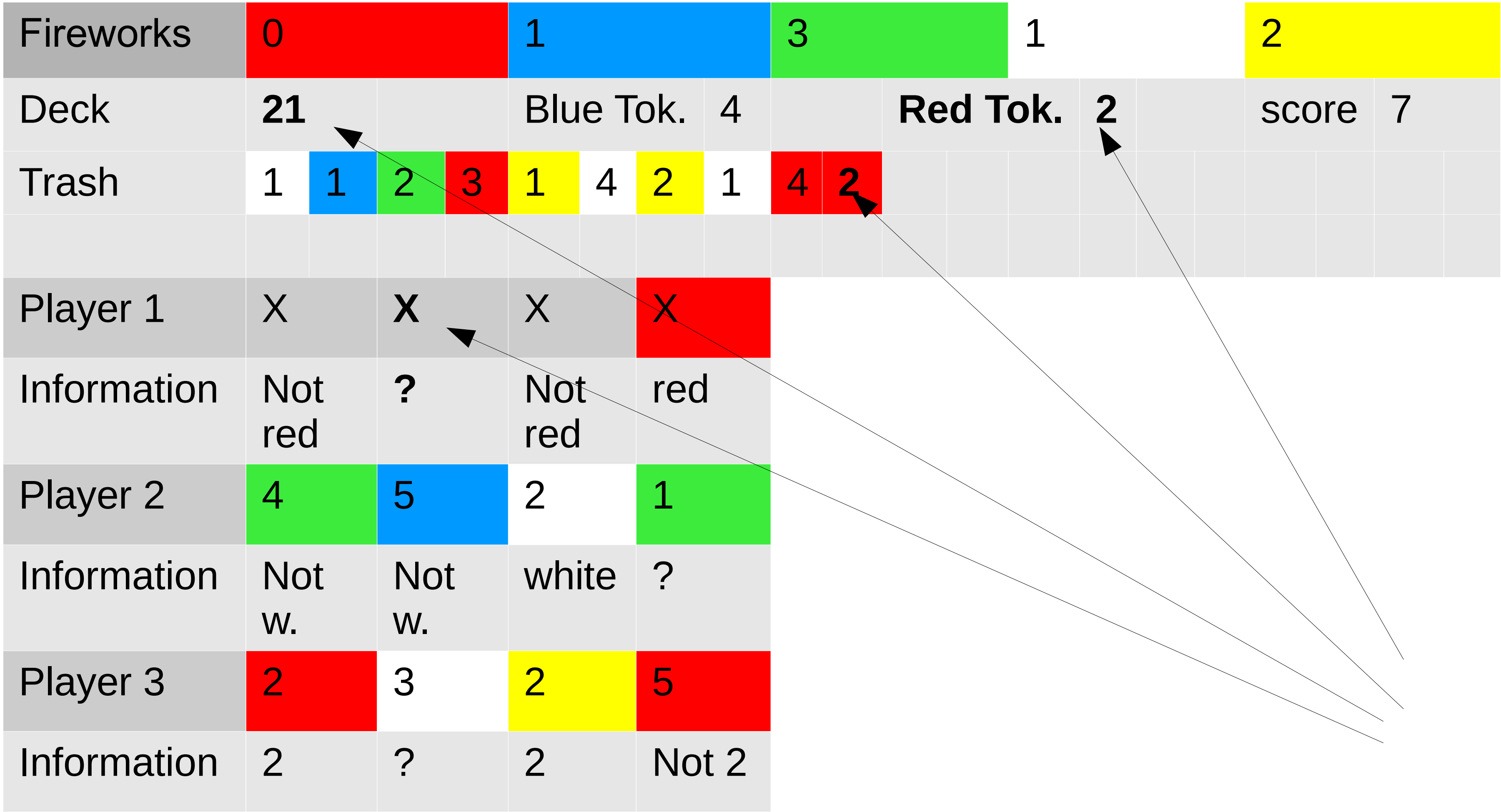
NP=3 players, NCPP=4 cards per player



Fireworks	0	1	3	1	2					
Deck	22		Blue Tok.	4		Red Tok.	3		score	7
Trash	1	1	2	3	1	4	2	1	4	
Player 1	X	**X**	X	X						
Information	Not red	red	Not red	red						
Player 2	4	5	2	1						
Information	Not w.	Not w.	white	?						
Player 3	2	3	2	5						
Information	2	?	2	Not 2						

Oops, it was red 2 ==> penalty

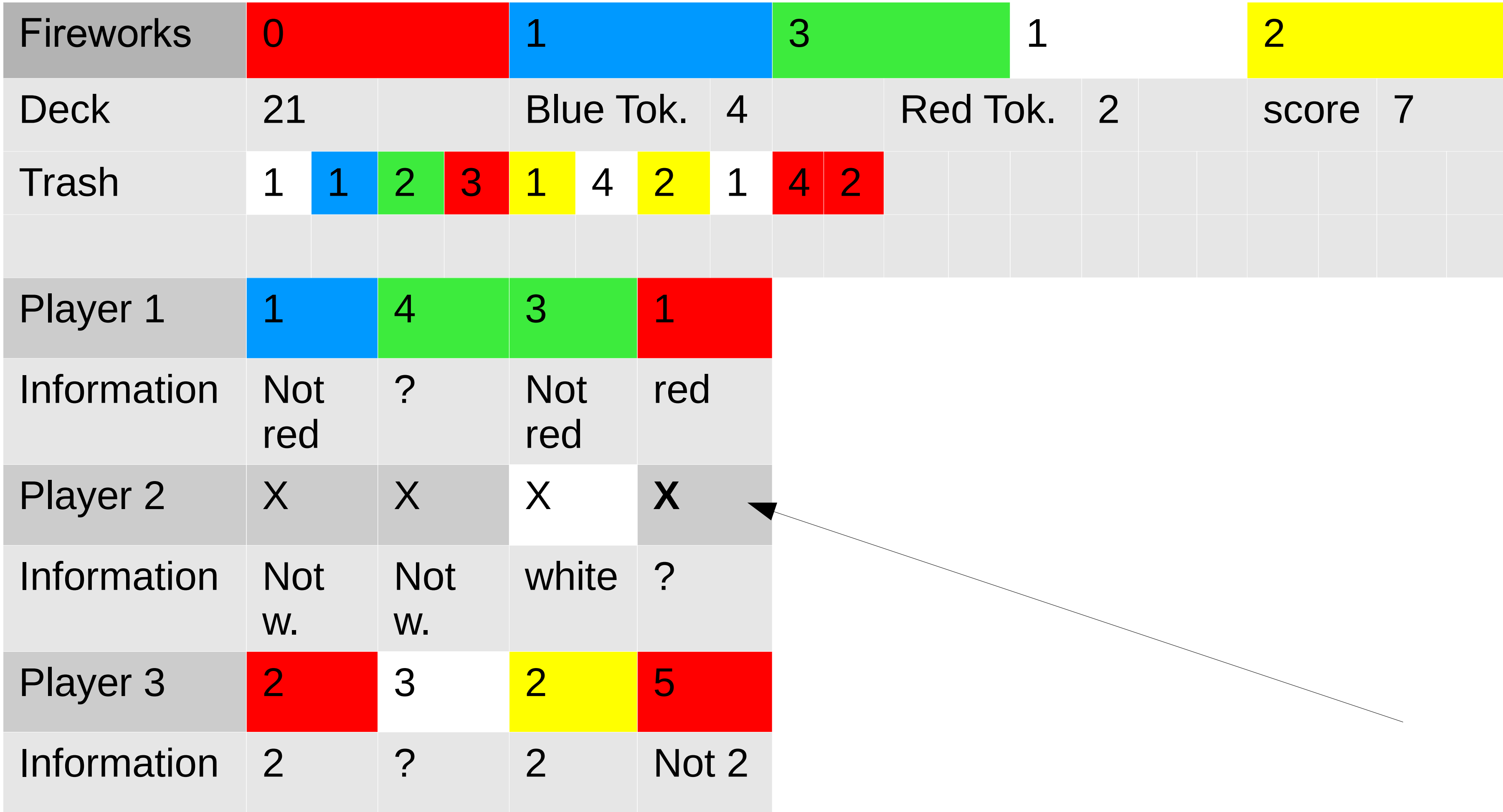
NP=3 players, NCPP=4 cards per player



Fireworks	🟥 0		🟦 1		🟩 3		1		🟨 2	
Deck	**21**		Blue Tok.	4		**Red Tok.**	**2**		score	7
Trash	1	🟦 1	🟩 2	🟥 3	🟨 1	4	🟨 2	1	🟥 4	🟥 **2**
Player 1	X	**X**	X	🟥 X						
Information	Not red	**?**	Not red	red						
Player 2	🟩 4	🟦 5	2	🟩 1						
Information	Not w.	Not w.	white	?						
Player 3	🟥 2	3	🟨 2	🟥 5						
Information	2	?	2	Not 2						

Player 2 to move

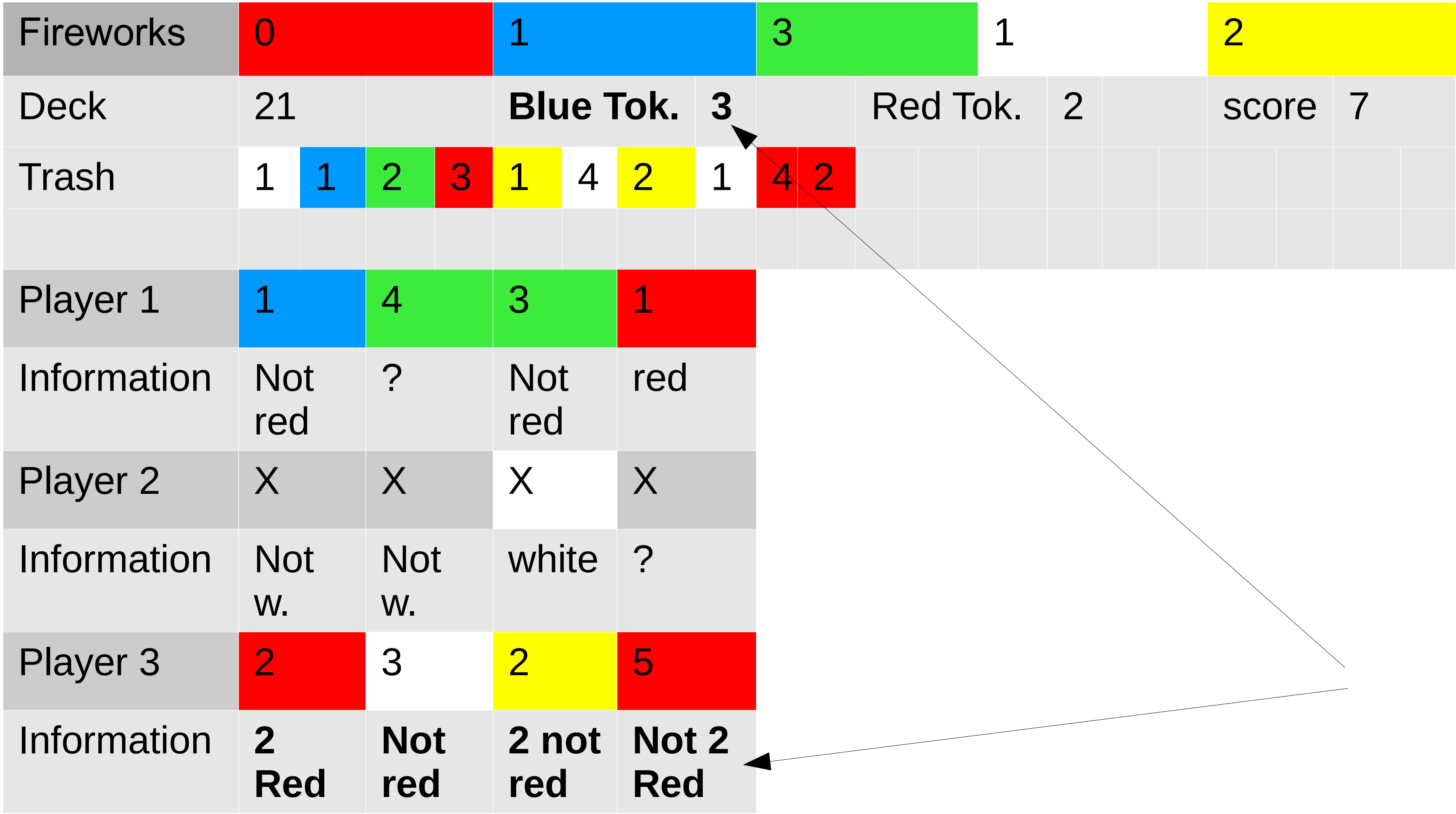
NP=3 players, NCPP=4 cards per player



Fireworks	0	1	3	1	2					
Deck	21	Blue Tok.	4	Red Tok.	2	score	7			
Trash	1	1	2	3	1	4	2	1	4	2
Player 1	1	4	3	1						
Information	Not red	?	Not red	red						
Player 2	X	X	X	**X**						
Information	Not w.	Not w.	white	?						
Player 3	2	3	2	5						
Information	2	?	2	Not 2						

P2 informs p3 with color = red

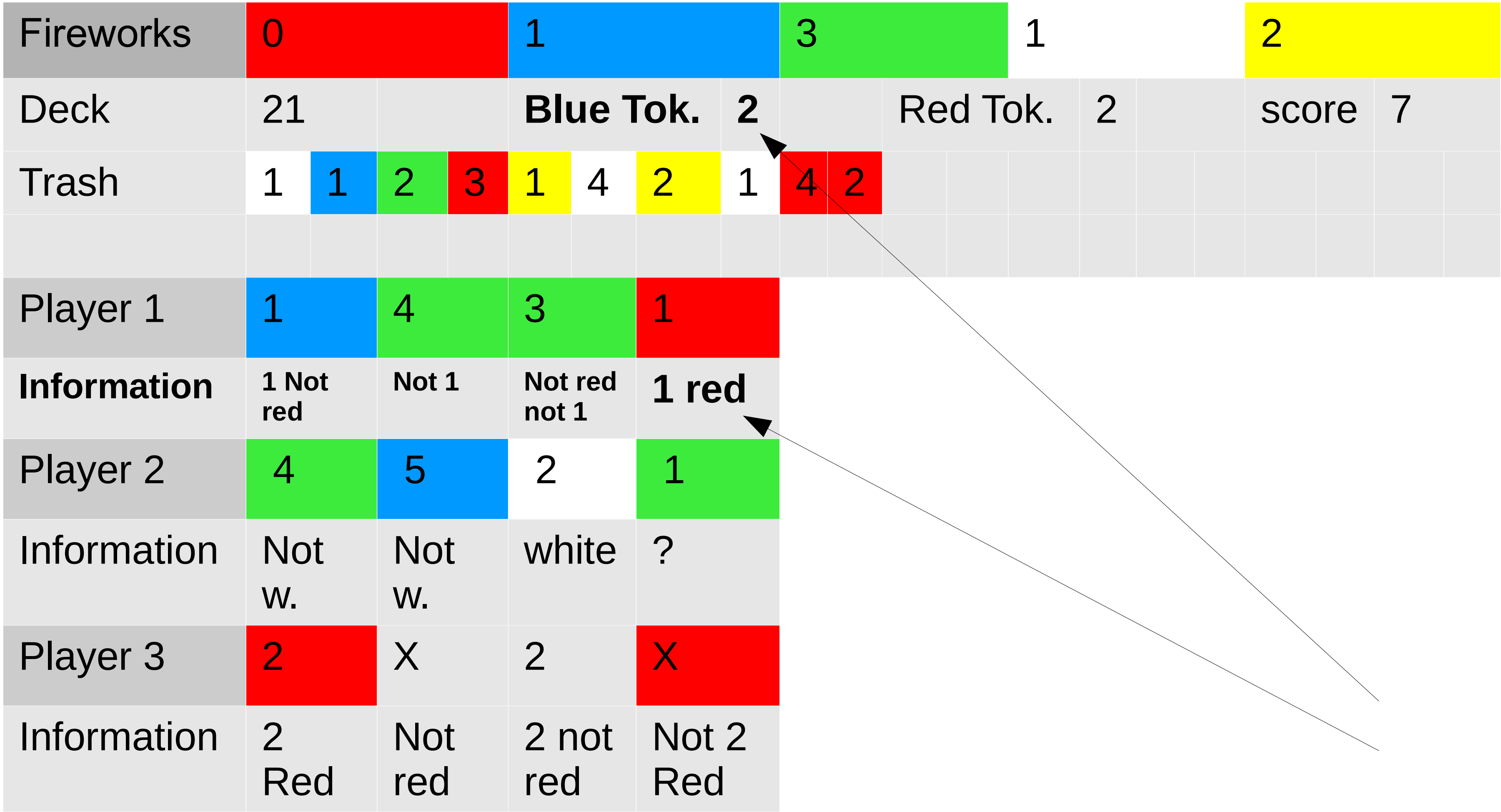
NP=3 players, NCPP=4 cards per player



Fireworks	🟥 0	🟦 1	🟩 3	⬜ 1	🟨 2					
Deck	21		**Blue Tok.**	**3**		Red Tok.	2		score	7
Trash	⬜ 1	🟦 1	🟩 2	🟥 3	🟨 1	⬜ 4	🟨 2	⬜ 1	🟥 4	🟥 2
Player 1	🟦 1	🟩 4	🟩 3	🟥 1						
Information	Not red	?	Not red	red						
Player 2	X	X	⬜ X	X						
Information	Not w.	Not w.	white	?						
Player 3	🟥 2	⬜ 3	🟨 2	🟥 5						
Information	**2 Red**	**Not red**	**2 not red**	**Not 2 Red**						

P3 informs p1 with height = 1

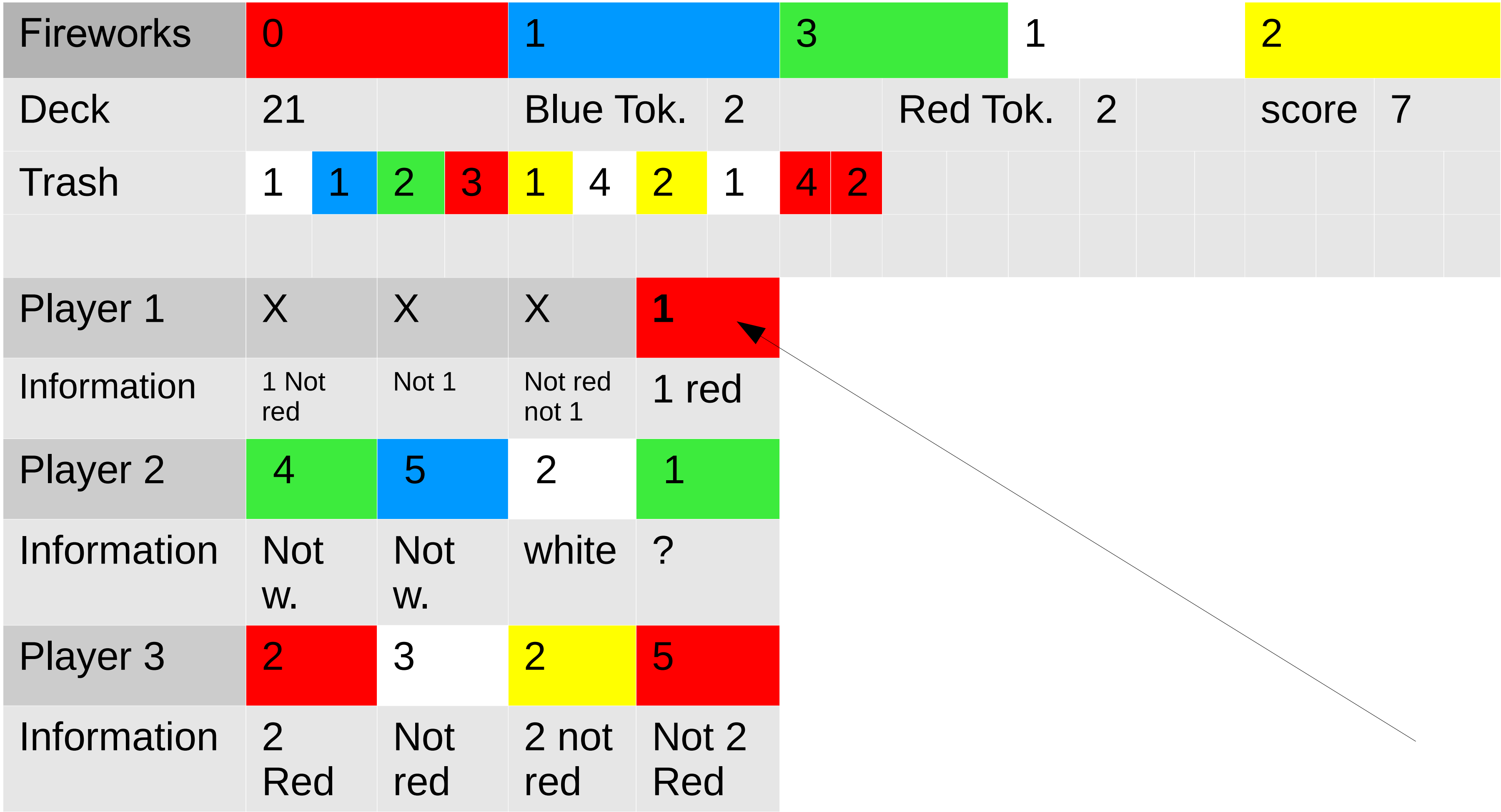
NP=3 players, NCPP=4 cards per player



Fireworks	🟥 0	🟦 1	🟩 3	⬜ 1	🟨 2					
Deck	21	**Blue Tok.**	**2**	Red Tok.	2	score	7			
Trash	⬜ 1	🟦 1	🟩 2	🟥 3	🟨 1	⬜ 4	🟨 2	⬜ 1	🟥 4	🟥 2
Player 1	🟦 1	🟩 4	🟩 3	🟥 1						
Information	**1 Not red**	**Not 1**	**Not red not 1**	**1 red**						
Player 2	🟩 4	🟦 5	⬜ 2	🟩 1						
Information	Not w.	Not w.	white	?						
Player 3	🟥 2	X	2	🟥 X						
Information	2 Red	Not red	2 not red	Not 2 Red						

P1 chooses to play card 4

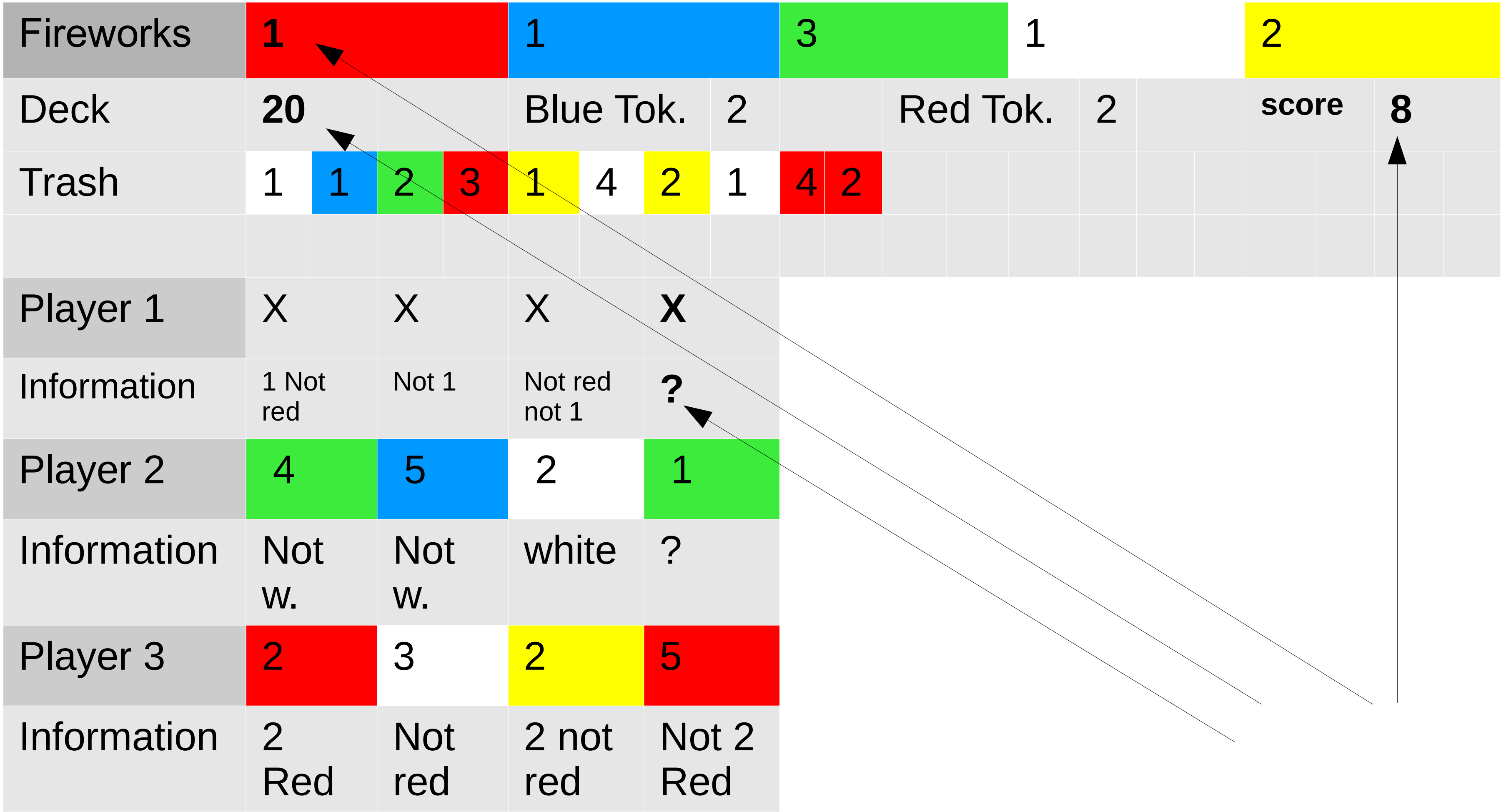
NP=3 players, NCPP=4 cards per player



Fireworks	🟥 0	🟦 1	🟩 3	⬜ 1	🟨 2					
Deck	21	Blue Tok.	2	Red Tok.	2	score	7			
Trash	⬜ 1	🟦 1	🟩 2	🟥 3	🟨 1	⬜ 4	🟨 2	⬜ 1	🟥 4	🟥 2
Player 1	X	X	X	🟥 **1**						
Information	1 Not red	Not 1	Not red not 1	1 red						
Player 2	🟩 4	🟦 5	⬜ 2	🟩 1						
Information	Not w.	Not w.	white	?						
Player 3	🟥 2	⬜ 3	🟨 2	🟥 5						
Information	2 Red	Not red	2 not red	Not 2 Red						

Success !

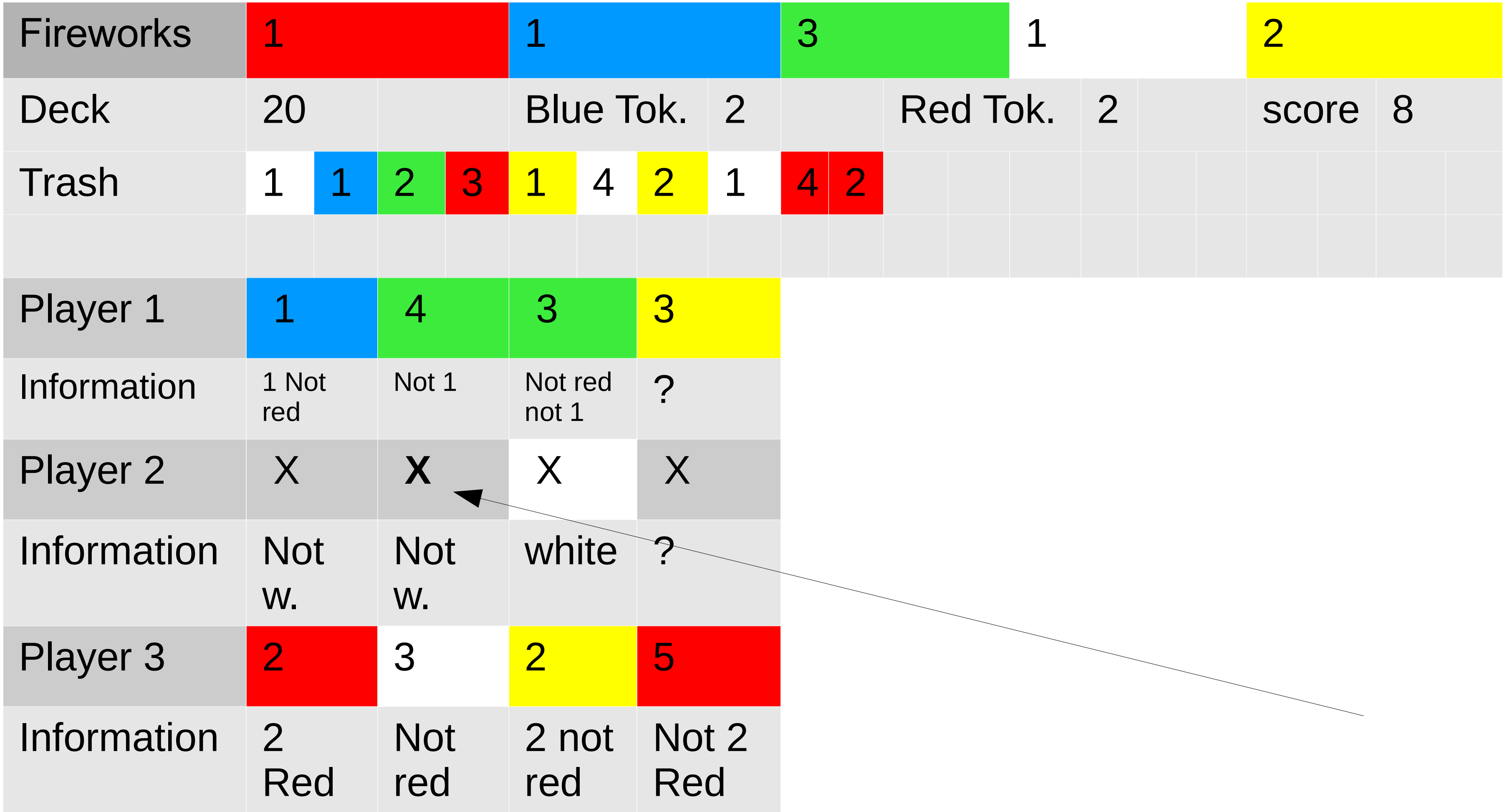
NP=3 players, NCPP=4 cards per player



Fireworks	🟥 **1**	🟦 1	🟩 3	⬜ 1	🟨 2					
Deck	**20**	Blue Tok.	2	Red Tok.	2	**score**	**8**			
Trash	⬜ 1	🟦 1	🟩 2	🟥 3	🟨 1	⬜ 4	🟨 2	⬜ 1	🟥 4	🟥 2
Player 1	X	X	X	**X**						
Information	1 Not red	Not 1	Not red not 1	**?**						
Player 2	🟩 4	🟦 5	⬜ 2	🟩 1						
Information	Not w.	Not w.	white	?						
Player 3	🟥 2	⬜ 3	🟨 2	🟥 5						
Information	2 Red	Not red	2 not red	Not 2 Red						

Player 2 chooses to discard card 2

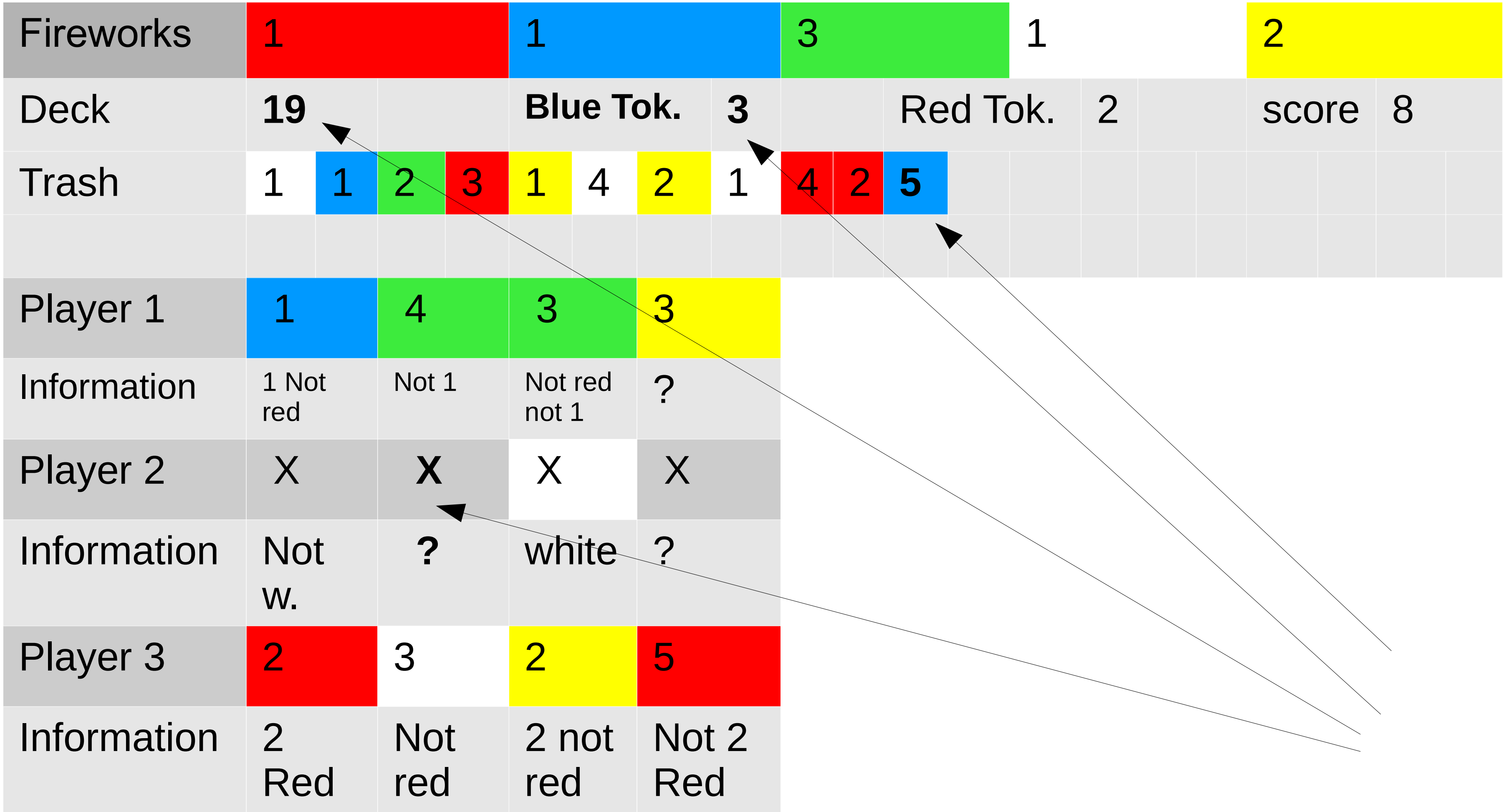
NP=3 players, NCPP=4 cards per player



Fireworks	🟥 1	🟦 1	🟩 3	⬜ 1	🟨 2					
Deck	20	Blue Tok.	2	Red Tok.	2	score	8			
Trash	⬜ 1	🟦 1	🟩 2	🟥 3	🟨 1	⬜ 4	🟨 2	⬜ 1	🟥 4	🟥 2
Player 1	🟦 1	🟩 4	🟩 3	🟨 3						
Information	1 Not red	Not 1	Not red not 1	?						
Player 2	X	**X**	⬜ X	X						
Information	Not w.	Not w.	white	?						
Player 3	🟥 2	⬜ 3	🟨 2	🟥 5						
Information	2 Red	Not red	2 not red	Not 2 Red						

One blue token is added

NP=3 players, NCPP=4 cards per player



Fireworks	🟥 1	🟦 1	🟩 3	⬜ 1	🟨 2						
Deck	**19**		**Blue Tok.**	**3**	Red Tok.	2	score	8			
Trash	⬜ 1	🟦 1	🟩 2	🟥 3	🟨 1	⬜ 4	🟨 2	⬜ 1	🟥 4	🟥 2	🟦 **5**
Player 1	🟦 1	🟩 4	🟩 3	🟨 3							
Information	1 Not red	Not 1	Not red not 1	?							
Player 2	X	**X**	⬜ X	X							
Information	Not w.	**?**	white	?							
Player 3	🟥 2	⬜ 3	🟨 2	🟥 5							
Information	2 Red	Not red	2 not red	Not 2 Red							

Hanabi

- The game ends when
  - The number of red tokens is zero
  - The score is 25
  - Each player has played once since the deck is empty

- Main features
  - Cooperative, N players
  - Hidden information
  - Finite episode
  - Explicit information moves


1
5
4
2
3
3

Previous work

- Osawa 2015
  - Partner models, NP=2, NCPP=5, <score> = 15.9
- Kosters, Van den berghe 2016
  - Miscellaneous, NP=3, NCPP=5, <score> = 15.4
- Franz 2016
  - MCTS, NP=4, NCPP=5, <score> = 17
- Cox 2015
  - Hat principle, NP=5, NCPP=4, <score> = 24.5

Goal

- Implement the hat principle
- Improve Cox's result
- Generalize to other NP and NCPP values

The hat principle

- Recreational mathematics
- Each hand is represented with a number – a hat – computed with a public program P1
- Each information move emitted by player P corresponds to the sum of hats player P sees. (public program P2)
- « Recommendation » version with NP=4
  - Each value of h corresponds to a « recommendation » (play card 1, play card 2, play card 3, play card 4, discard card 1, discard card 2, discard card 3, discard card 4)
  - 4 playing moves + 4 discarding moves = 8 moves
  - Hat h, such that 0 <= h <8
- Information move by player P :
  - Compute S(P) the sum of hats that P sees.
  - Perform the corresponding information move.
  - With a subtraction, players Q different from P can deduce their own hat.

The hat principle

- Number of information moves (NIM)
  - NIM = (NP-1) NIMP
  - NIMP = 10
    - 5 colors + 5 heights (many work)
  - NIMP = 2
    - Color or height (Cox’s work)
- Importance of the rule set
  - Informing a player with an empty set : allowed or not
  - NIM >= H

Allowing all information moves or not ?

Player	1	2	3	2

- Wikipedia and many sources including our work
  - No forbidden information moves
  - NIMP = 10

- Cox 2015
  - No corresponding card in the player’s hand ==> forbidden information moves
    - Color = Green
    - Color = Yellow
    - Height = 4
    - Height = 5
  - NIMP = 2

- Commercial ruleset
  - Not mentioned (!)

The hat principle

- « Information » version
  - A hat -> a (subset of) value(s) of a card
  - Each hand has a specific card to be informed
    - Given by a public program P3
      - (Highest playing probability,
      - Left most non informed card)
- Ruleset
  - If NIM >= 25 : Each hat -> unique card value
  - Otherwise a hat may correspond to a subset of card values
- Effect
  - A player is quickly informed with its cards’ values.
  - As if the players could see their own cards

Artificial players

- Certainty player
  - Play or disgard totally informed cards only
- Confidence player
  - Without proof of the contrary, assumes an informed card is playable
- Seer player
  - Sees its own card but not the deck
- Hat players
  - Recommendation player
  - Information player
- Depth-one tree search player
  - Use an above player as a policy in a depth-one Monte-Carlo search
  - Uses NCD plausible card distributions
  - (Kuhn 1955) polynomial time assignment problem algorithm

Experiments

- Team made up with NP copies of the same player
- Test set
  - NG games (each with one card distribution)
  - NG = 100 for tree search players
  - NG = 10,000 for knowledge-based players
- « near-optimality » :
  - approaching the seer empirical score on a given test set.
- Settings
  - 3 Ghz , 10 minutes / game at most
  - No memory issue
  - NCD = 1, 10, 100, 1k, 10k.

Results (knowledge based players)

Certainty (Cert), Confidence (Conf), Hat recommendation (Hrec) and Hat information (Hinf)
For NP = 2, 3, 4, 5 ; NCPP = 3, 4, 5 ; NG = 10,000

NP		Cert			Conf			Hrec			Hinf	
	3	4	5	3	4	5	3	4	5	3	4	5
2	10.3	10.7	11.1	16.9	16.7	15.8	15.8	16.9	17.8	5.9	6.4	6.7
3	12.9	13.0	13.5	19.4	19.2	17.9	22.8	23.8	23.8	18.7	19.4	18.9
4	14.4	14.7	14.1	20.3	19.7	17.9	23.2	23.5	22.8	24.3	24.6	24.4
5	15.2	14.4	12.8	20.6	19.2	16.8	23.2	22.6	21.0	24.6	24.7	24.3

Hat information, NP=5 NCPP=4, histogram of scores, NG = 10,000

Score	19	20	21	22	23	24	25
%	0.01	0.05	0.17	1.19	3.62	13.66	81.30

Results (depth-one tree search players)

Tree search players using :
Confidence (Conf), Hat recommendation (Hrec), Hat information (Hinf), Seer
For NP = 2, 3, 4, 5 ;   NCPP = 3, 4, 5 ; NG = 100 ;   NCD = 100, 1k, 10k

NP		Conf			Hrec			Hinf			Seer	
	3	4	5	3	4	5	3	4	5	3	4	5
2	19.2	19.4	19.0	16.40	17.38	18.53				23.10	24.46	24.91
3	20.7	21.1	20.4	23.96	24.56	24.70				24.62	24.97	25.00
4	21.5	21.0	19.7	24.34	24.60	24.45	24.72	24.96	24.91	24.91	25.00	24.99
5	22.0	20.4	18.0	24.26	24.30	22.68	24.85	24.92	24.76	24.96	24.98	24.96

Tree search + Hat information, NP=5 NCPP=4, Histogram of scores , NG = 100

Score	19	20	21	22	23	24	25
%	0	0	0	0	0	8	92


1
5
4
2
3

Conclusions and future work

- Summary
  - The hat convention « kills » the game for the computer
  - Current work :
    - Small step upward with depth-one search
      - 75 % → 90 % for NP=5 and NCPP=4
    - Generalization to NP = 2, 3, 4, 5 and NCPP = 3, 4, 5
    - Use of the hungarian method of the card assignment problem

- Future work :
  - Neural Network approach :
  - Is it possible to learn the hat convention ?
    - In self play ? What is the architecture ?
    - With a teacher ?

Thank you for your attention!

Questions ?

bruno.bouzy@parisdescartes.fr