

21<sup>st</sup> Canadian Multidisciplinary Road Safety Conference  
May 8-11, 2011, Halifax, Nova Scotia

# Using Data Mining Techniques to Understand Collision Processes

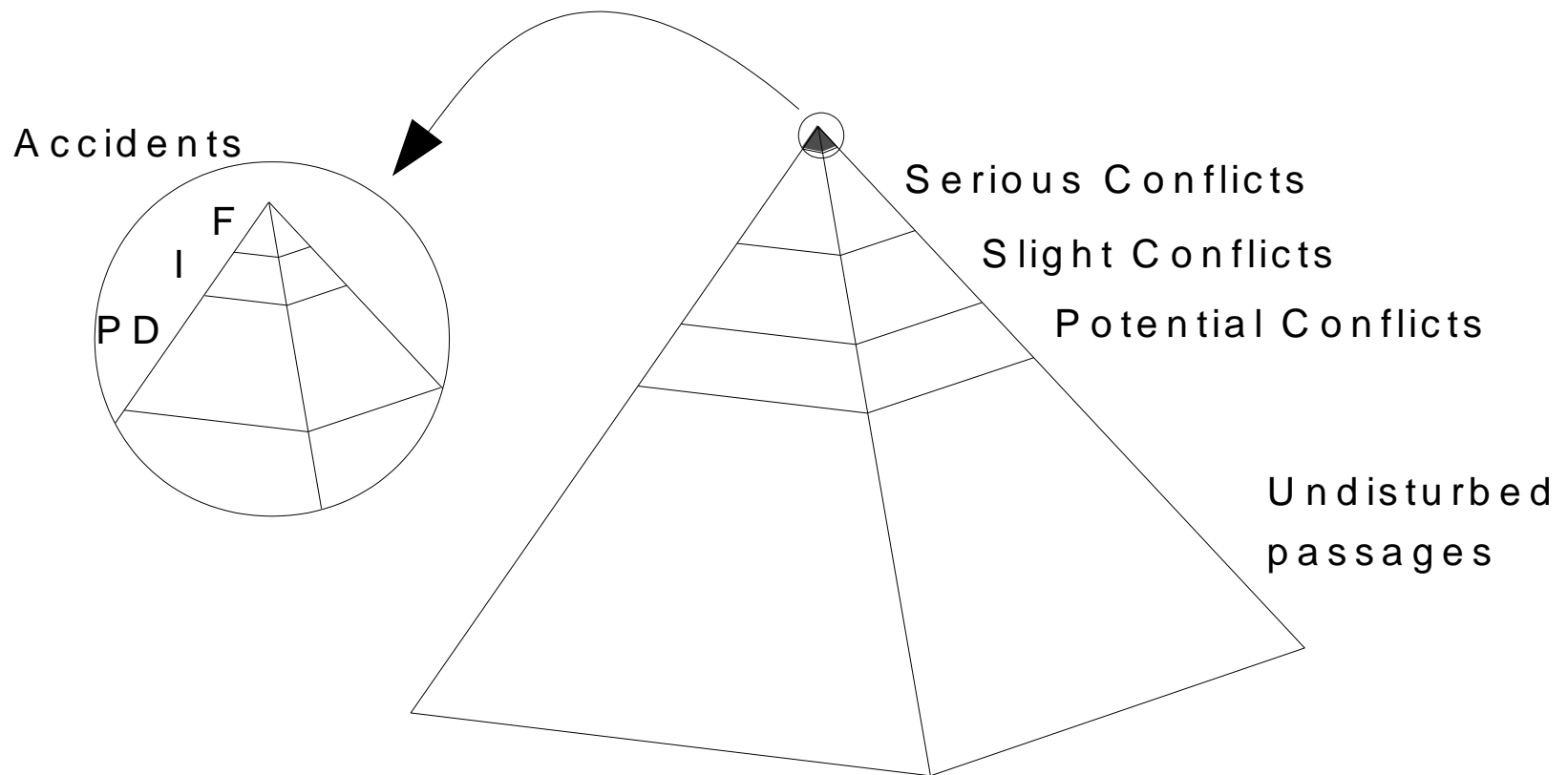
Nicolas Saunier, Nadia Mourji and Bruno Agard  
École Polytechnique de Montréal



# Motivation

- Need for surrogate measures of road safety
- Difficult validation of surrogate measures of safety, debates about conflicts, definitions...

# Hypothesis: the Safety Hierarchy



# Objective

- Understand collision processes to
  - design better counter-measures
  - develop better surrogate measures based on better-known relationships between interactions with and without a collision
- How?
  - continuous traffic data collection: record all traffic events, e.g. using video sensors
  - Knowledge Discovery and Data Mining (KDD) techniques

# Trajectories Extraction from Video Data



Conflicts

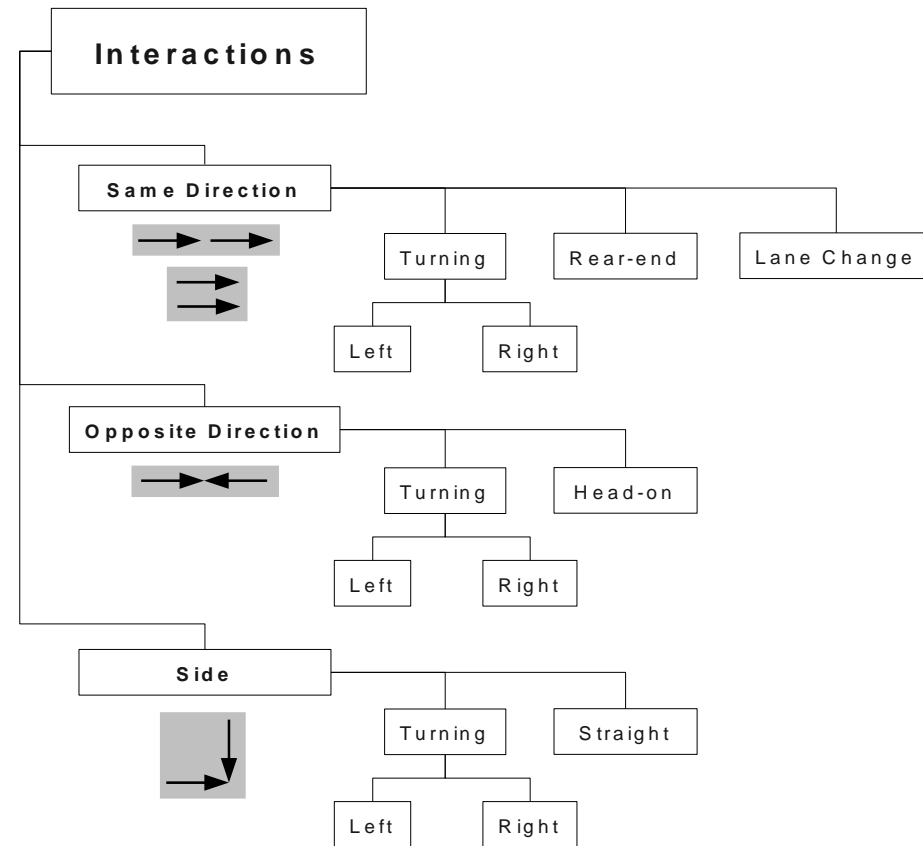


Collisions

*(Saunier, Sayed and Ismail 2010)*

# Interaction Description: Categorical Attributes

Categorical Attributes	Values
Type of day	weekday, week end
Lighting condition	daytime, twilight, nighttime
Weather condition	normal, rain, snow
Interaction category	see figure →
Interaction outcome	conflict, collision

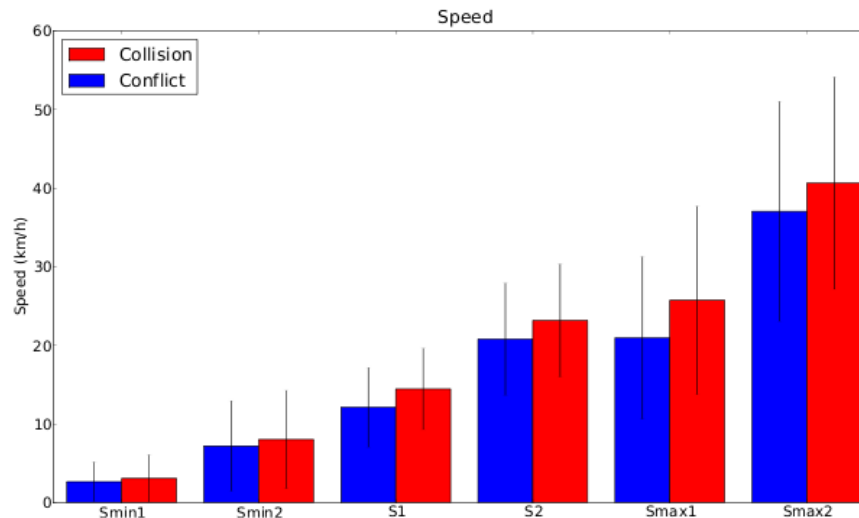
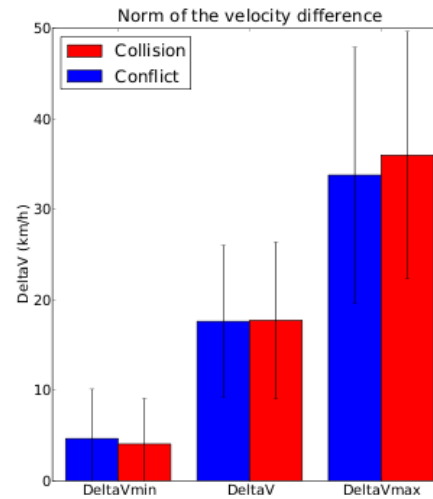


# Interaction Description: Numerical Attributes

Numerical Attributes	Units
Road user type passenger car van, 4x4, SUV bus...	number of road users per type
Road user origin...	number of road users per origin
Type of evasive action No evasive action Braking Swerving Acceleration	number of evasive actions per evasive action
3 attributes from the speed differential $\Delta v$ (min, max and mean)	km/h
6 values from the road users' speeds (min, max and mean for each)	km/h

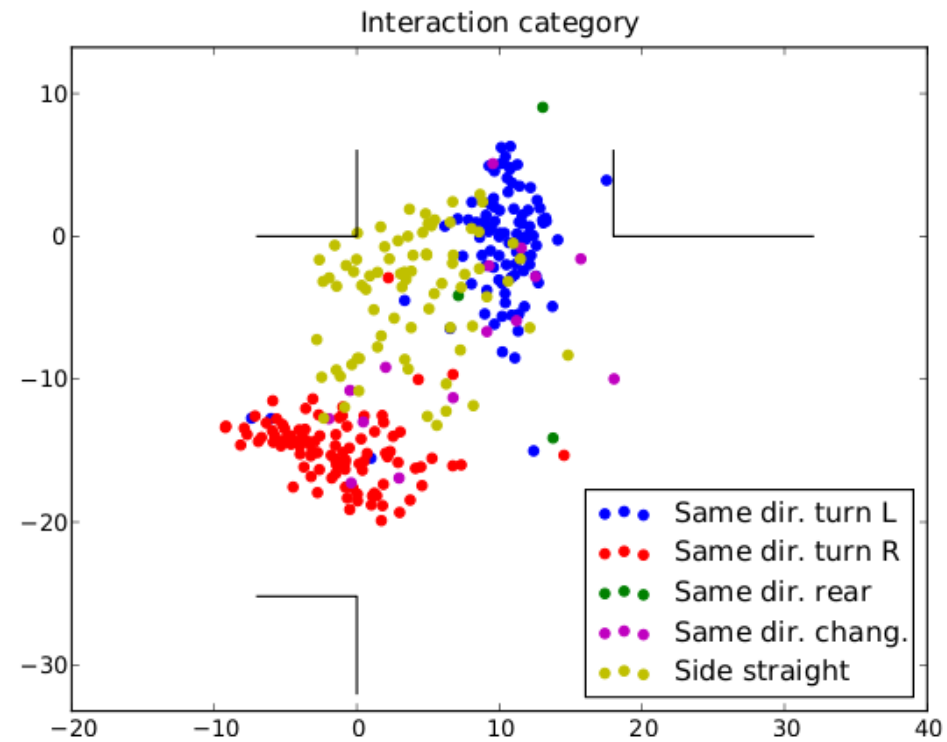
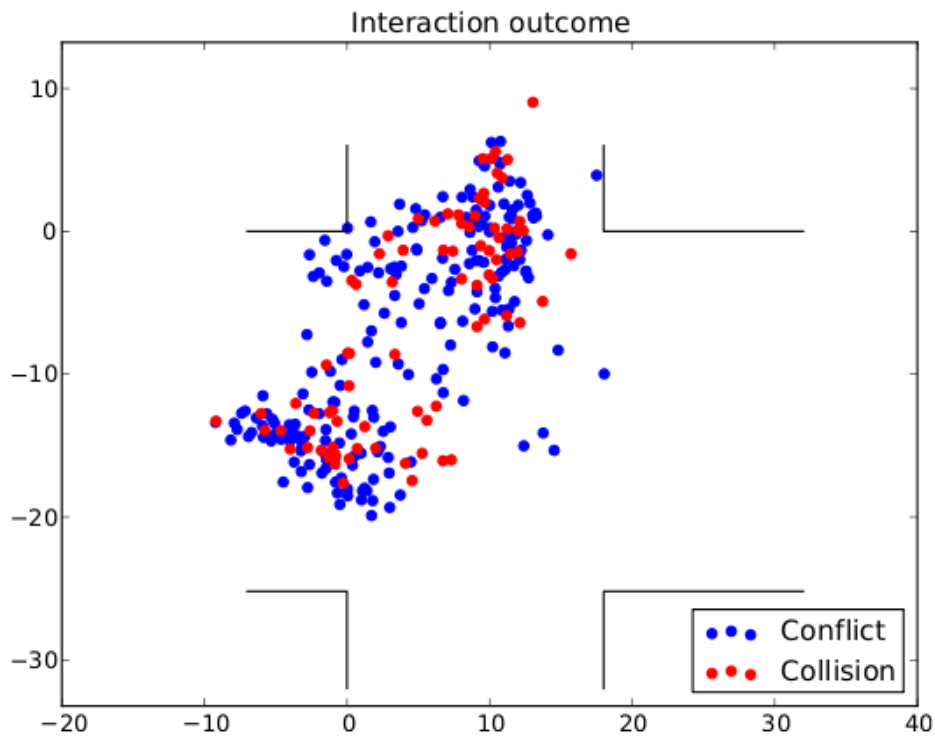
Coarse symmetric description of the relative road users' trajectories

# Descriptive Analysis



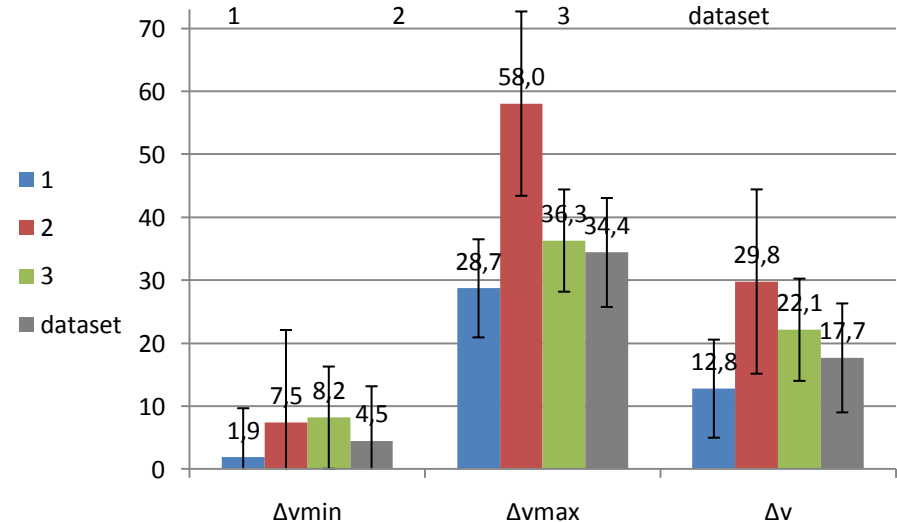
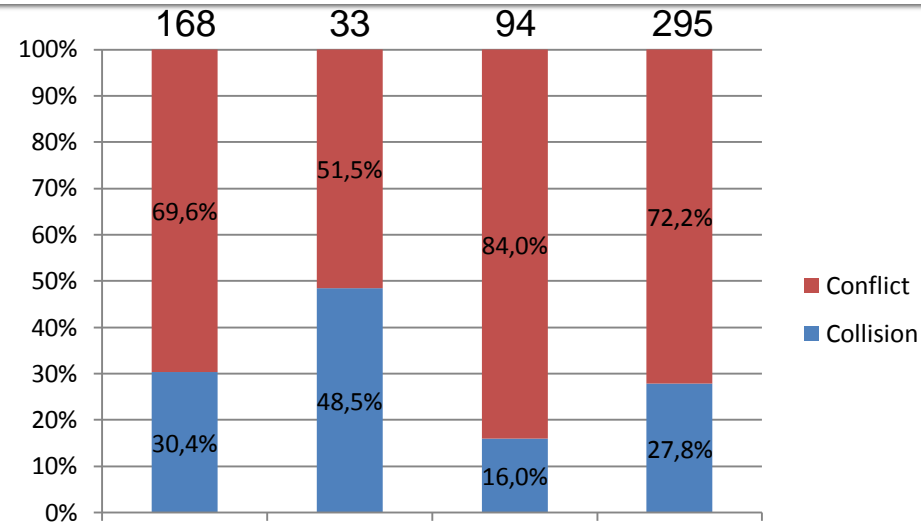
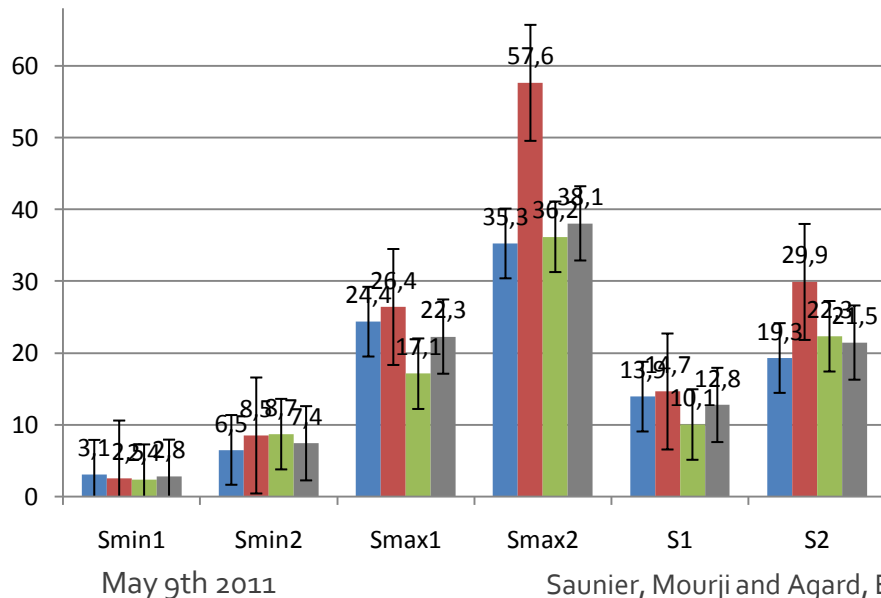


# Descriptive Analysis

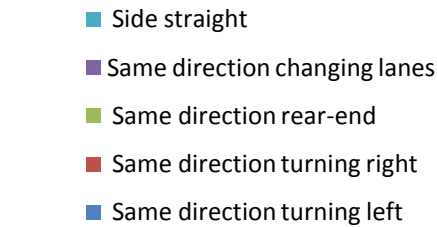
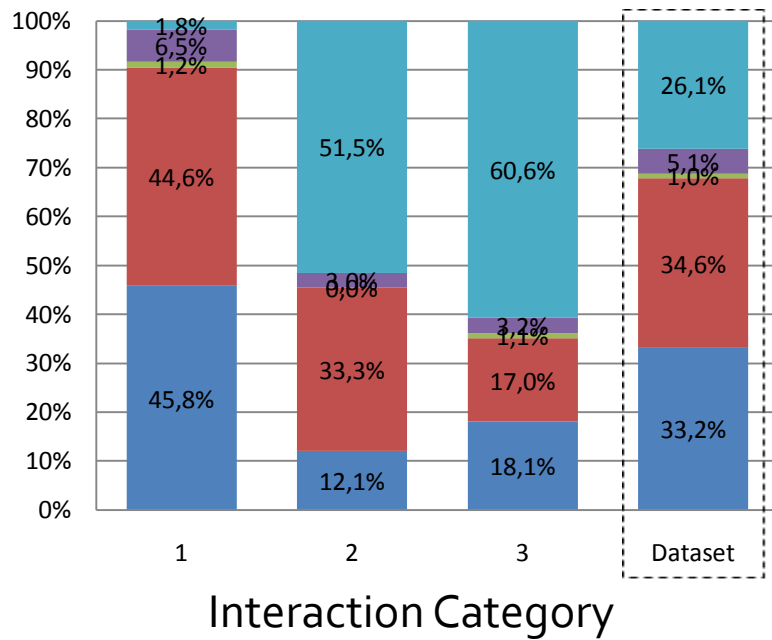


# Clustering Results

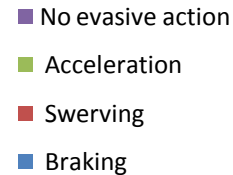
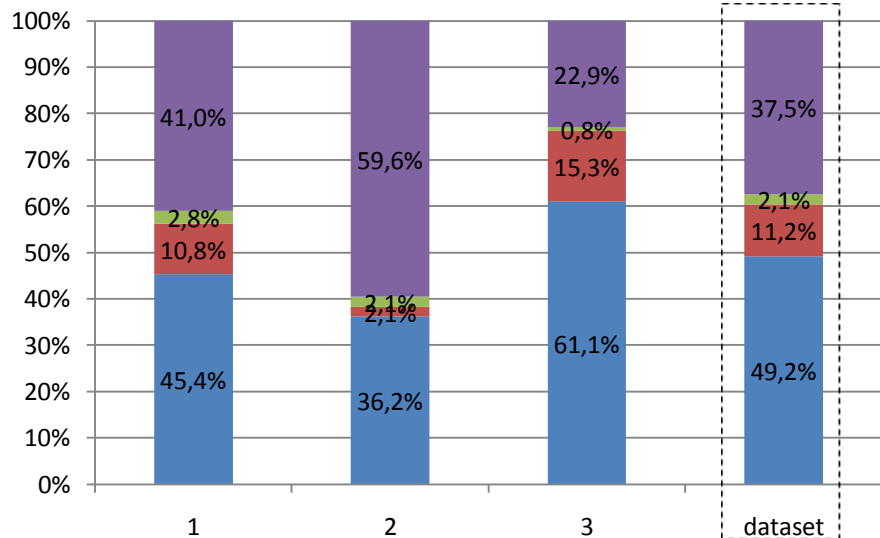
- k-medoid algorithm
- 3 groups
- using all attributes (except origins and outcome)
- Euclidean distance, with specific distance for interaction categories



# Clustering Results



## Evasive Action



# Clustering Results

	<b>CLUSTER 1</b>	<b>CLUSTER 2</b>	<b>CLUSTER 3</b>
<b>NUMBER OF INTERACTIONS</b>	168	33	94
<b>SPEED DIFFERENTIALS</b>	Lowest speed differentials	Highest speed differentials	Medium speed differentials
<b>SPEEDS</b>	Lowest to medium speeds alternating with cluster 3	Highest speeds	Lowest to medium speeds alternating with cluster 1
<b>INTERACTION OUTCOME</b>	30.4 % of collisions 79.6 % of conflicts	48.5 % of collisions 51.5 % of conflicts	16.0 % of collisions 84.0 % of conflicts
<b>INTERACTION CATEGORY</b>	45.8 % Same direction turning left 44.6 % Same direction turning right	51.5 % Side straight 33.3 % Same direction turning right	60.6 % Side Straight 18.0 % Same direction turning left 17.0 % Same direction turning right
<b>TYPE OF ROAD USERS</b>	59.7 % Passenger car 30.9 % 4X4, VAN, VUS 8.6 % Truck	55.4 % Passenger car 44.6 % 4X4, VAN, VUS	53.4 % Passenger car 41.1 % 4X4, VAN, VUS 5.5 % Truck
<b>TYPE OF EVASIVE ACTIONS</b>	41.0 % No evasive action 45.4 % Braking	59.6 % No evasive action 36.2 % Braking	22.9 % No evasive action 61.1 % Braking 15.3 % Swerving
<b>TYPE OF DAY</b>	59.5 % Weekday 40.5 % Week-end	30.3 % Weekday 69.7 % Week-end	78.7 % Weekday 21.3 % Week-end

# Logit Model of Interaction Outcome

	Coefficient	Std. Error	z-stat	Slope
const	-1.72947	1.28607	-1.3448	
Same direction turning left	2.78372	1.04016	2.6763	0.439349
Same direction turning right	1.72514	1.0261	1.6813	0.244256
Side straight	4.44196	1.34845	3.2941	0.757887
Braking	-4.1418	0.571796	-7.2435	-0.701337
Swerving	-2.67496	0.767919	-3.4834	-0.17601
No evasive action	1.41745	0.546812	2.5922	0.160854
$\Delta v$	-0.180444	0.0553516	-3.2600	-0.0208568
$s_2$	0.138837	0.0504446	2.7523	0.0160476

Coefficient of determination  $R^2$ : 0.5462  
 Correct prediction rate: 90.2 %

# Conclusion

- Method to understand collision processes
  - find groups of similar conflicts and collisions
  - supplementary evidence that not all conflicts should be used as surrogates for all collisions
- Work in progress:
  - compare the whole time series of interaction description variables
  - collect large datasets of trajectories
- Open science: share data and code (open source)

# Questions?

Contact

nicolas.saunier@polymtl.ca

More on

<http://nicolas.saunier.confins.net>