Graduiertenzentrum. OWL
OWL Graduate Center

Invited Talk on Wednesday, 13th July, 2016
Lemgo, Germany

Fraunhofer
IOSB-INA

Towards interactive Machine Learning for solving complex problems

Andreas Holzinger

Holzinger Group, Research Unit HCI-KDD, Institute for Medical Informatics, Statistics and Documentation, Medical University Graz

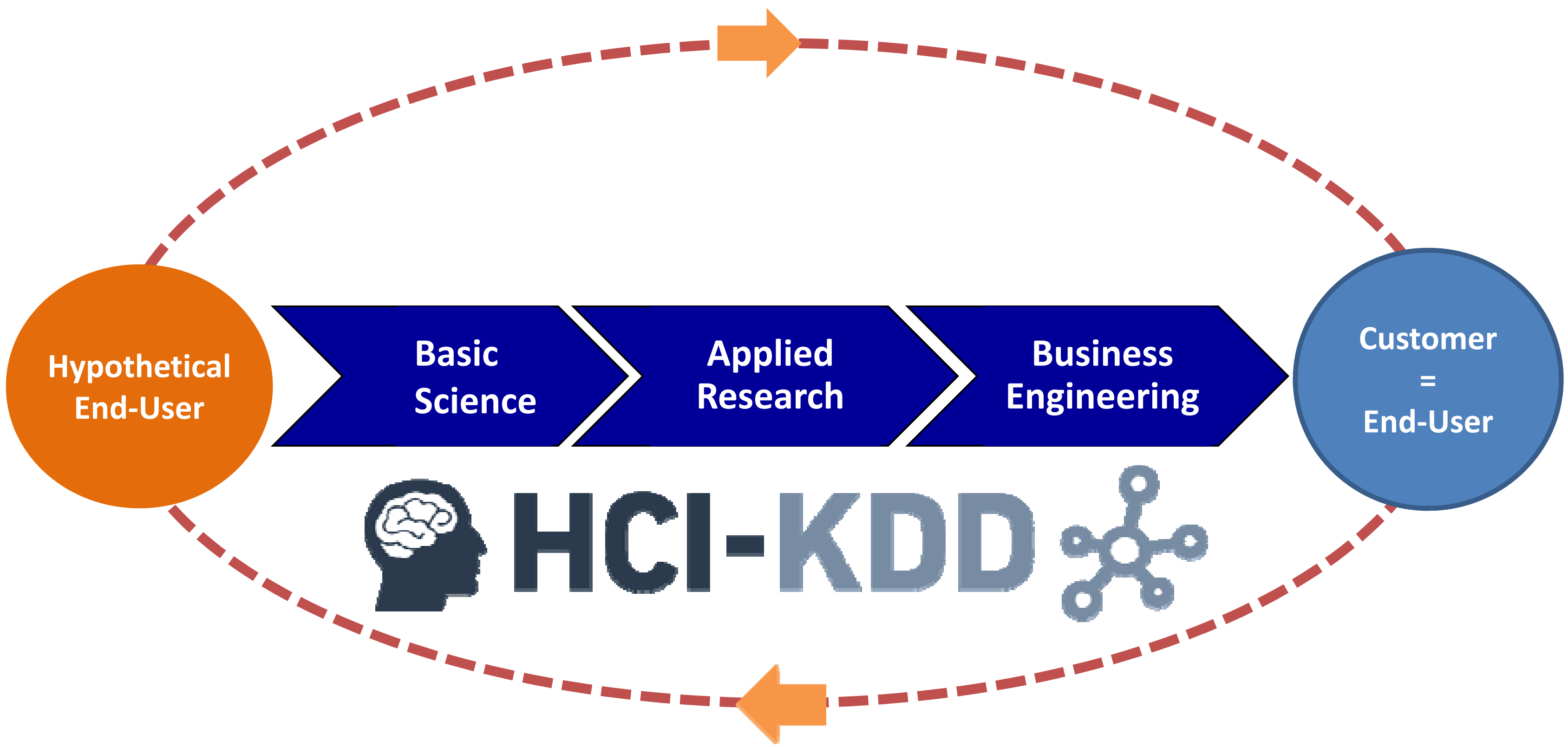
&

Institute of Information Systems and Computer Media, Graz University of Technology

- The HCI-KDD approach
- ML and Health Informatics
- ML state-of-the-art
- aML versus iML
- A few examples of iML
- Future Outlook

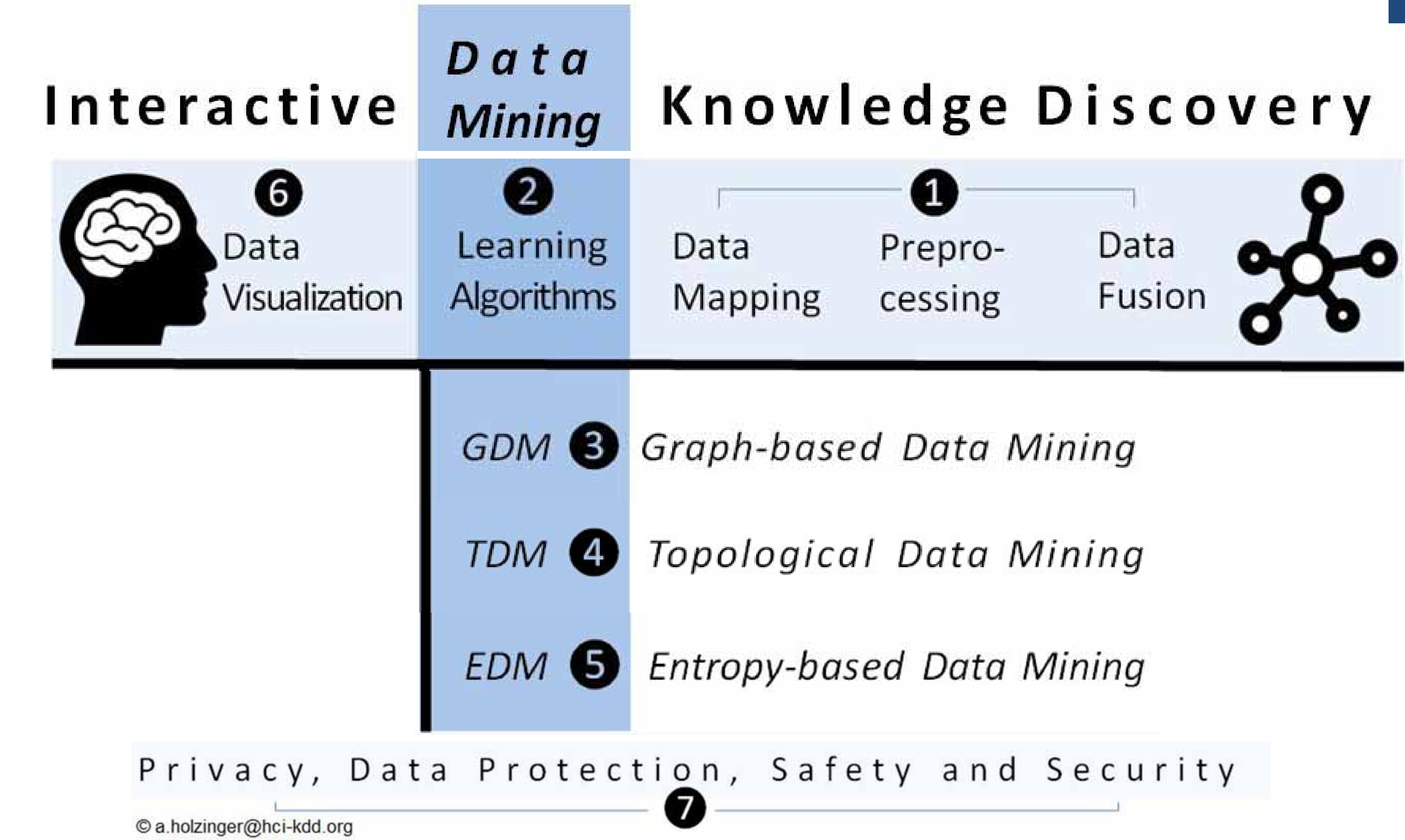

HCI-KDD
You
http://hci-kdd.org/international-expert-network

Science is testing crazy ideas – Engineering is putting these ideas into Business



Holzinger, A. 2011. Successful Management of Research and Development, Norderstedt: BoD.

ML needs a concerted effort fostering integrated research

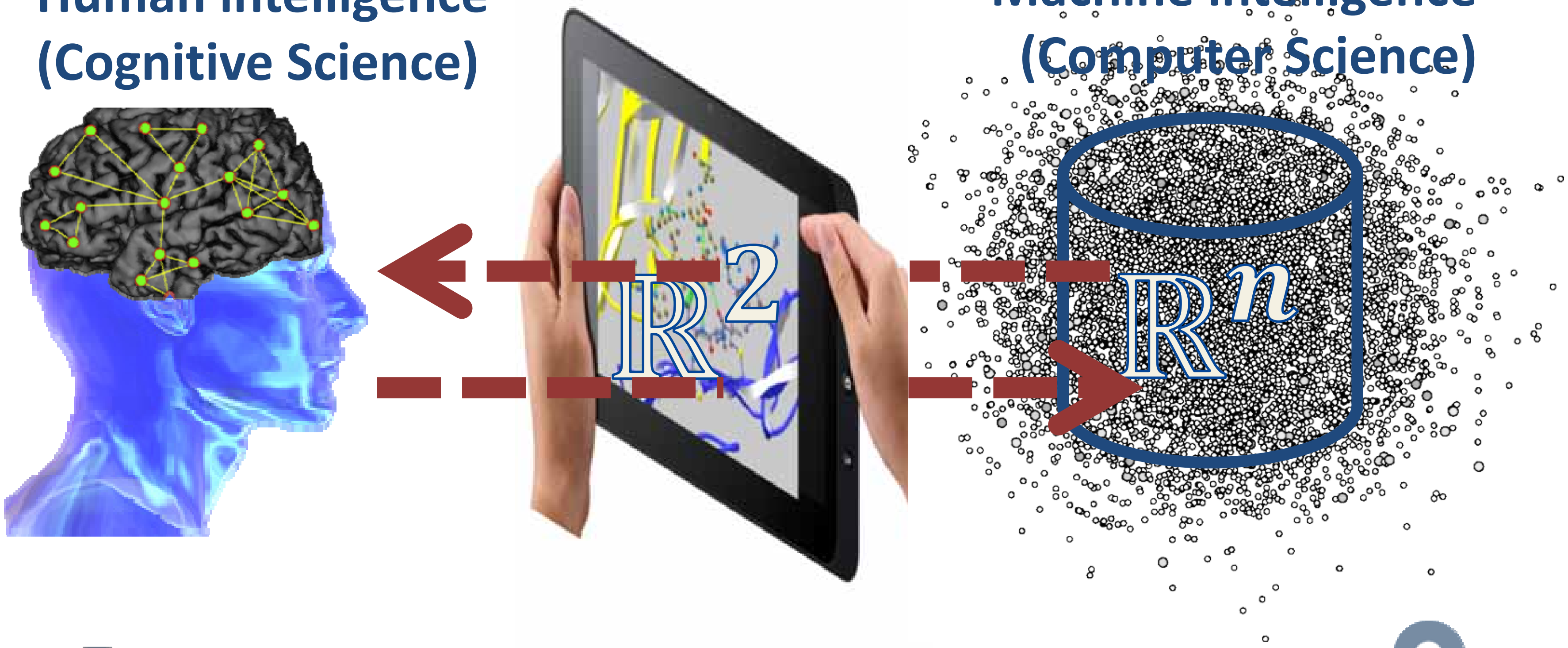


Holzinger, A. 2014. Trends in Interactive Knowledge Discovery for Personalized Medicine: Cognitive Science meets Machine Learning. IEEE Intelligent Informatics Bulletin, 15, (1), 6-14.

Grand challenge: Transfer results from $\mathbb{R}^n$ to $\mathbb{R}^2$

Human intelligence (Cognitive Science)

Machine intelligence (Computer Science)



Holzinger, A. (2013). Human–Computer Interaction & Knowledge Discovery (HCI-KDD): What is the benefit of bringing those two fields to work together? In: Lecture Notes in Computer Science 8127 (pp. 319-328)

- CS aims to reverse engineer **human intelligence**;
- ML provides powerful sources of insight into *how* **machine intelligence** is possible.
- CS therefore raises challenges for, and draws inspiration from ML;
- Insights about the human mind may help inspire **new directions for ML …**

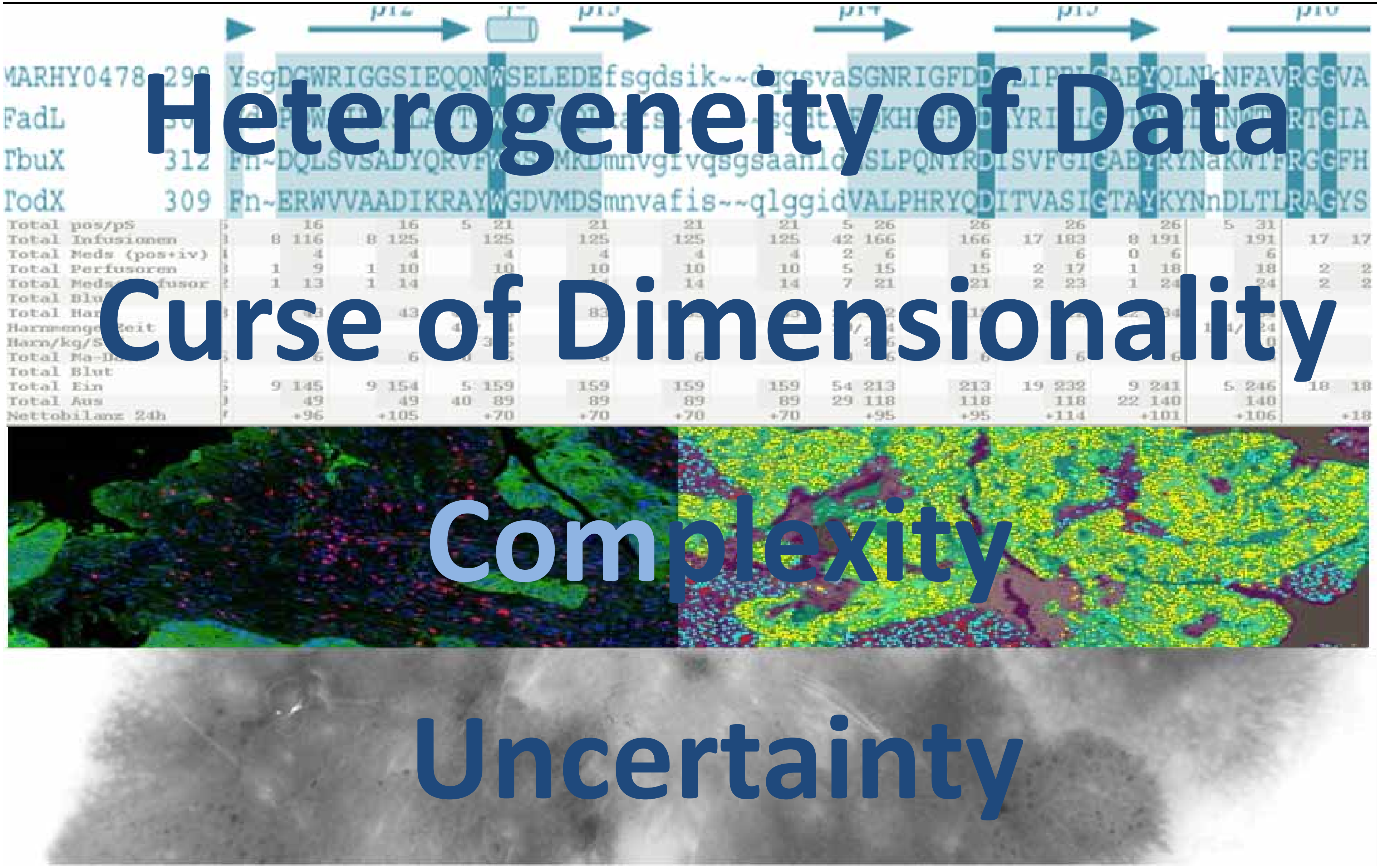
Application Area: Health Informatics

Why is this application area complex ?

Our central hypothesis:
Information may bridge this gap

Holzinger, A. & Simonic, K.-M. (eds.) 2011. *Information Quality in e-Health. Lecture Notes in Computer Science LNCS 7058, Heidelberg, Berlin, New York: Springer.*


Where is the
problem in building
this bridge?



Holzinger, A., Dehmer, M. & Jurisica, I. 2014. Knowledge Discovery and interactive Data Mining in Bioinformatics - State-of-the-Art, future challenges and research directions. BMC Bioinformatics, 15, (S6), I1.

Probabilistic Information p(x)

Of great help was …

Probabilistic Information p(x)

Thomas Bayes
1701 - 1761

Bayes, T. (1763). An Essay towards solving a Problem in the Doctrine of Chances (Postum communicated by Richard Price). Philosophical Transactions, 53, 370-418.

$$p(x_i) = \sum P(x_i, y_j)$$

$$p(x_i, y_j) = p(y_j|x_i)P(x_i)$$

Bayes' Rule is a corollary of the Sum Rule and Product Rule:

$$p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{\sum p(x_i, y_j)p(x_i)}$$

Barnard, G. A., & Bayes, T. (1958). Studies in the history of probability and statistics: IX. Thomas Bayes's essay towards solving a problem in the doctrine of chances. Biometrika, 45(3/4), 293-315.

d … *data* h … *hypothesis* H … $\{H_1, H_2, \dots, H_n\}$ $\forall\, h, d$ …

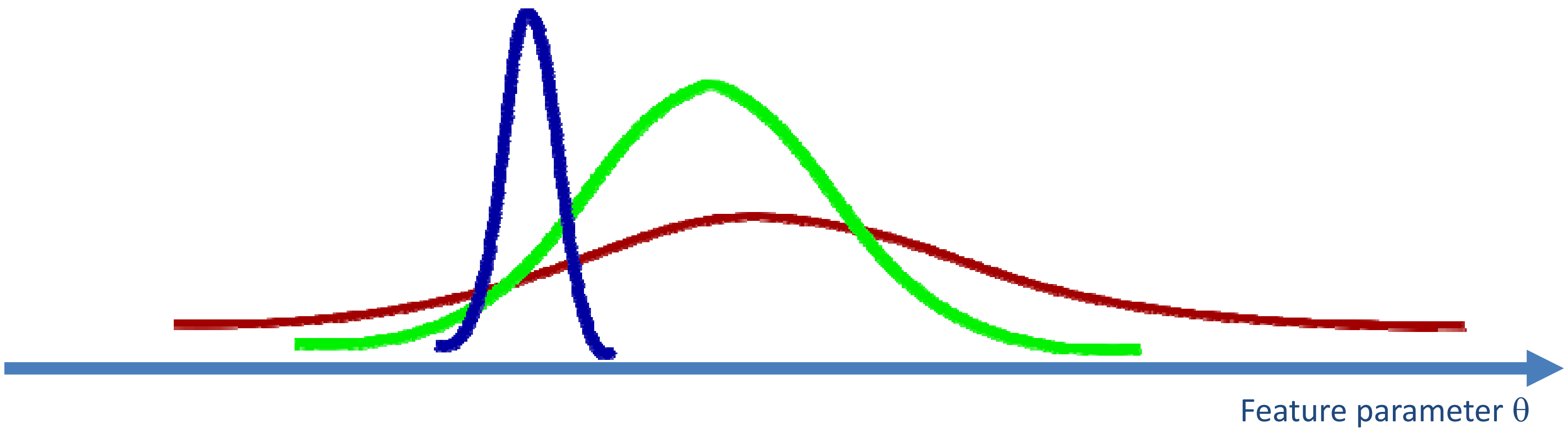
Likelihood

Prior Probability

$$p(h|d) = \frac{p(d|h) * p(h)}{\sum_{h \in H} p(d|h') \, p(h')}$$

Posterior Probability

Problem in $\mathbb{R}^n \rightarrow$ complex

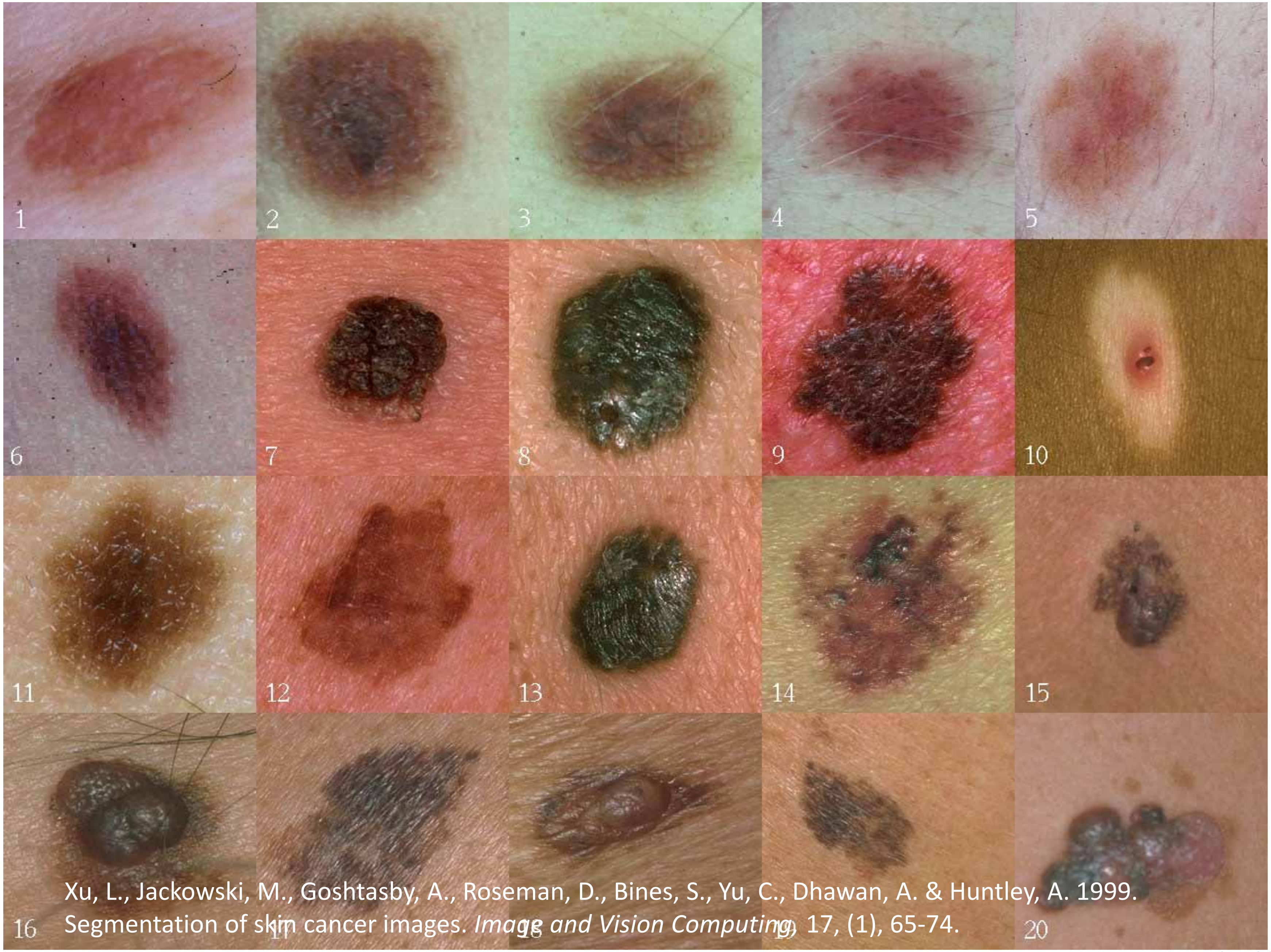


Biomedical Example

To find structural anomalies in such data …

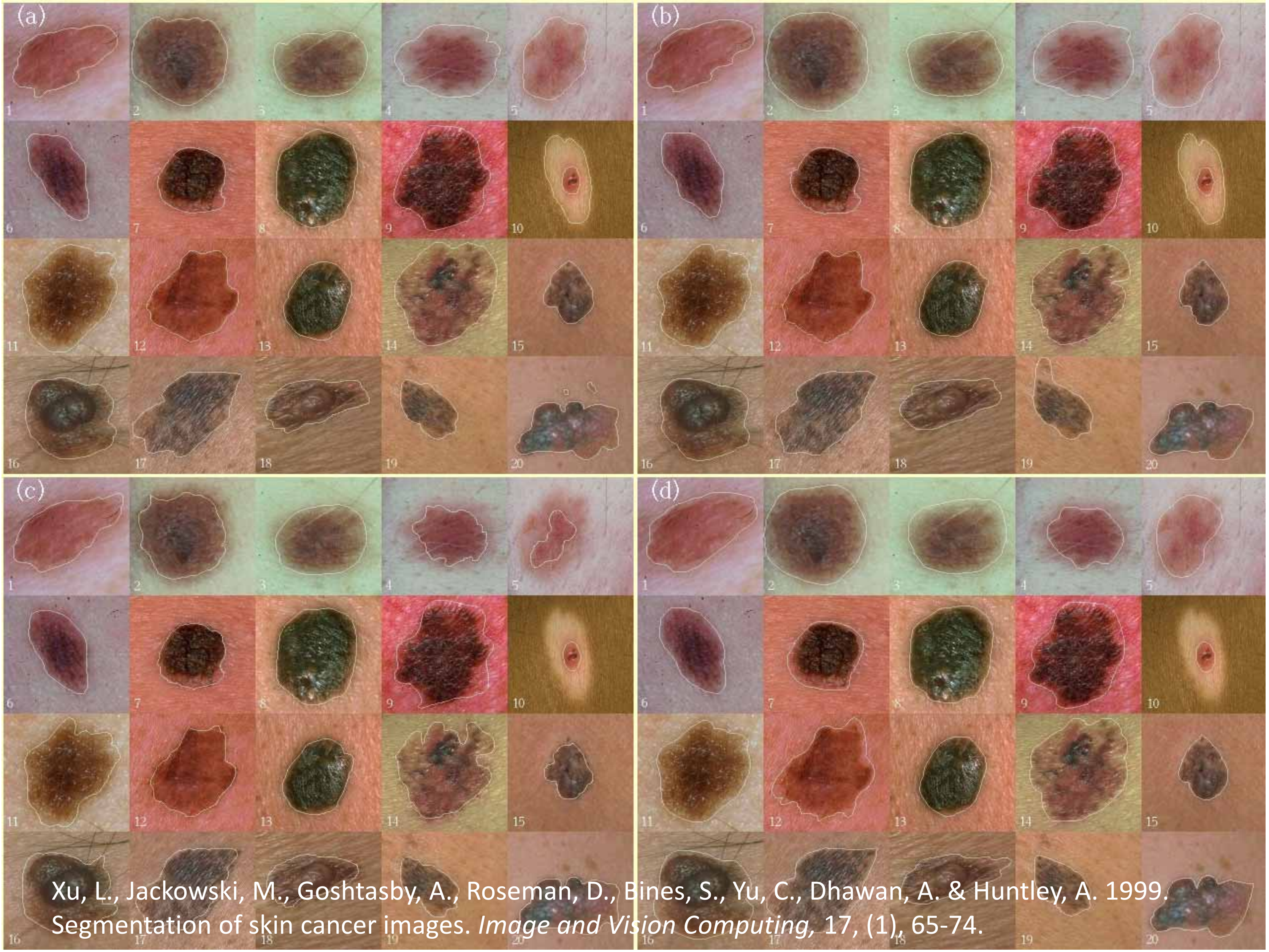
Holzinger, A., Malle, B., Bloice, M., Wiltgen, M., Ferri, M., Stanganelli, I. & Hofmann-Wellenhof, R. 2014. On the Generation of Point Cloud Data Sets: Step One in the Knowledge Discovery Process. *In: Lecture Notes in Computer Science, LNCS 8401.* Berlin Heidelberg: Springer, pp. 57-80, doi:10.1007/978-3-662-43968-5_4.

Learning from previous examples …



Xu, L., Jackowski, M., Goshtasby, A., Roseman, D., Bines, S., Yu, C., Dhawan, A. & Huntley, A. 1999. Segmentation of skin cancer images. *Image and Vision Computing*, 17, (1), 65-74.

The more examples we have the better …



Xu, L., Jackowski, M., Goshtasby, A., Roseman, D., Bines, S., Yu, C., Dhawan, A. & Huntley, A. 1999. Segmentation of skin cancer images. *Image and Vision Computing,* 17, (1), 65-74.

Big Data is good for automatic Machine Learning

$$\mathcal{D} = x_{1:n} = \{x_1, x_2, ..., x_n\}$$

$$p(\mathcal{D}|\theta)$$

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta) * p(\theta)}{p(\mathcal{D})}$$

$$posterior = \frac{likelihood * prior}{evidence}$$

The inverse probability allows to learn from data, infer unknowns, and make predictions

$$\max_{\mathbf{x} \in \mathcal{A} \subset \mathbb{R}^d} f(\mathbf{x})$$

$$p(h|d) \propto p(\mathcal{D}|\theta) * p(h)$$

$$p(f(x)|\mathcal{D}) \propto p(\mathcal{D}|f(x)) * p(f(x))$$

- Machine Learning is the development of algorithms which can **learn from data**
- assessment of **uncertainty,** making **predictions**
- **Automating automation - getting computers to program themselves – let the data do the work!**

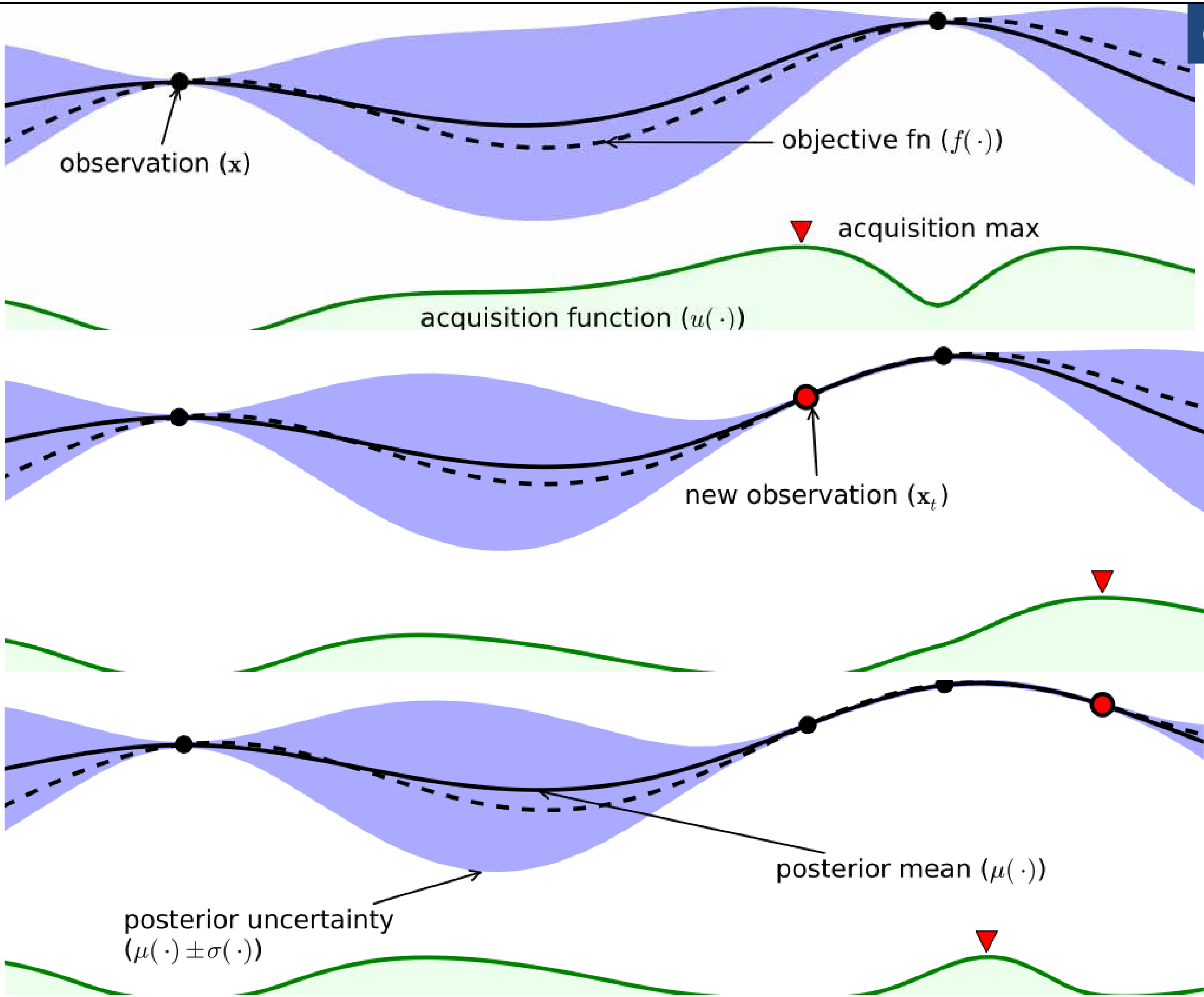
- **Newton, Leibniz, … developed calculus – mathematical language for describing and dealing with rates of change**
- **Bayes, Laplace, … developed probability theory - the mathematical language for describing and dealing with uncertainty**

Someone - born 152 km away from Lemgo – is important

Gaussian Processes

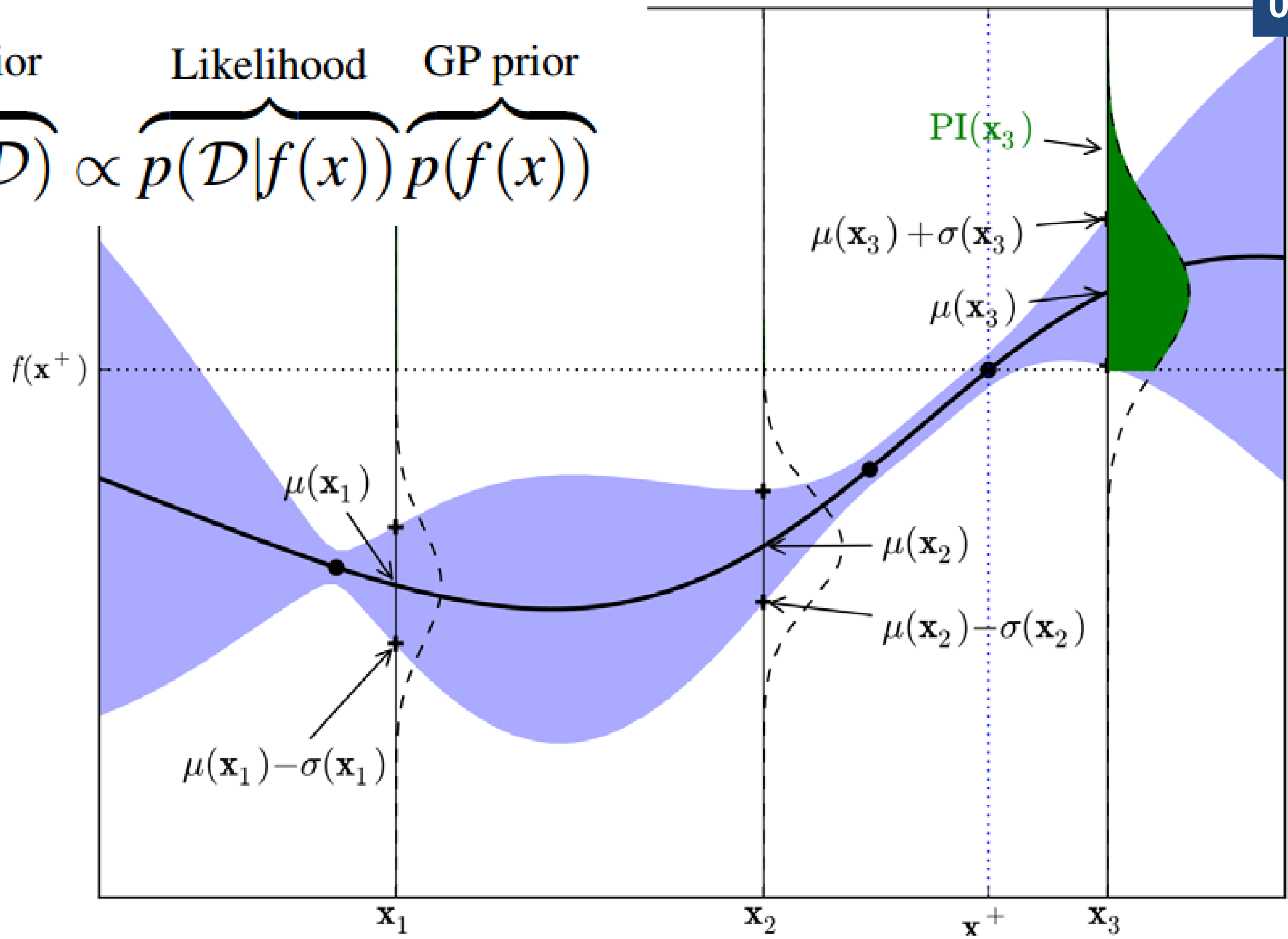
From Bayesian Optimization to Gaussian Process (GP) approximation

07



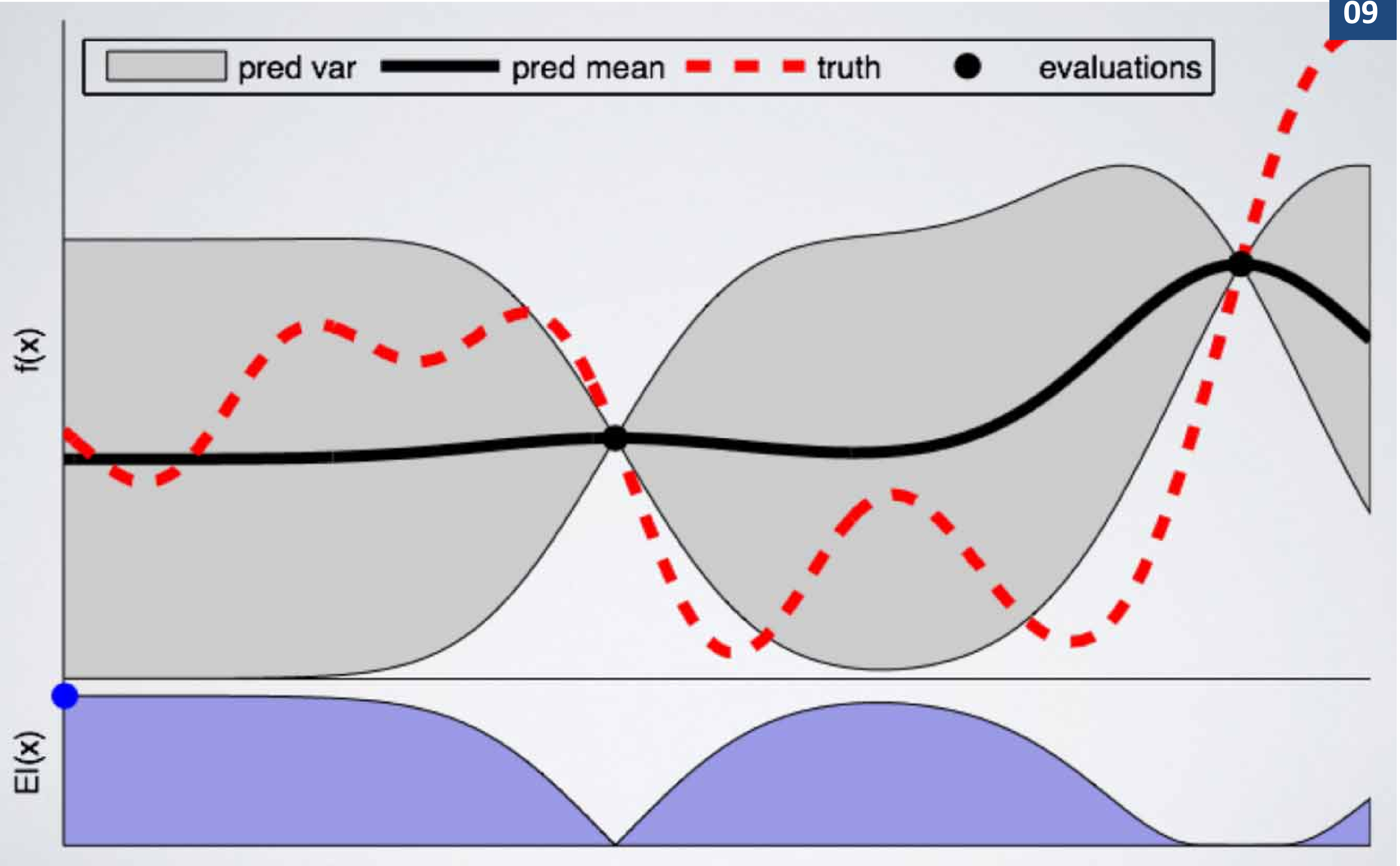
Brochu, E., Cora, V. M. & De Freitas, N. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning arXiv:1012.2599.

$$\underbrace{p(f(x)|\mathcal{D})}_{\text{GP posterior}} \propto \underbrace{p(\mathcal{D}|f(x))}_{\text{Likelihood}} \underbrace{p(f(x))}_{\text{GP prior}}$$

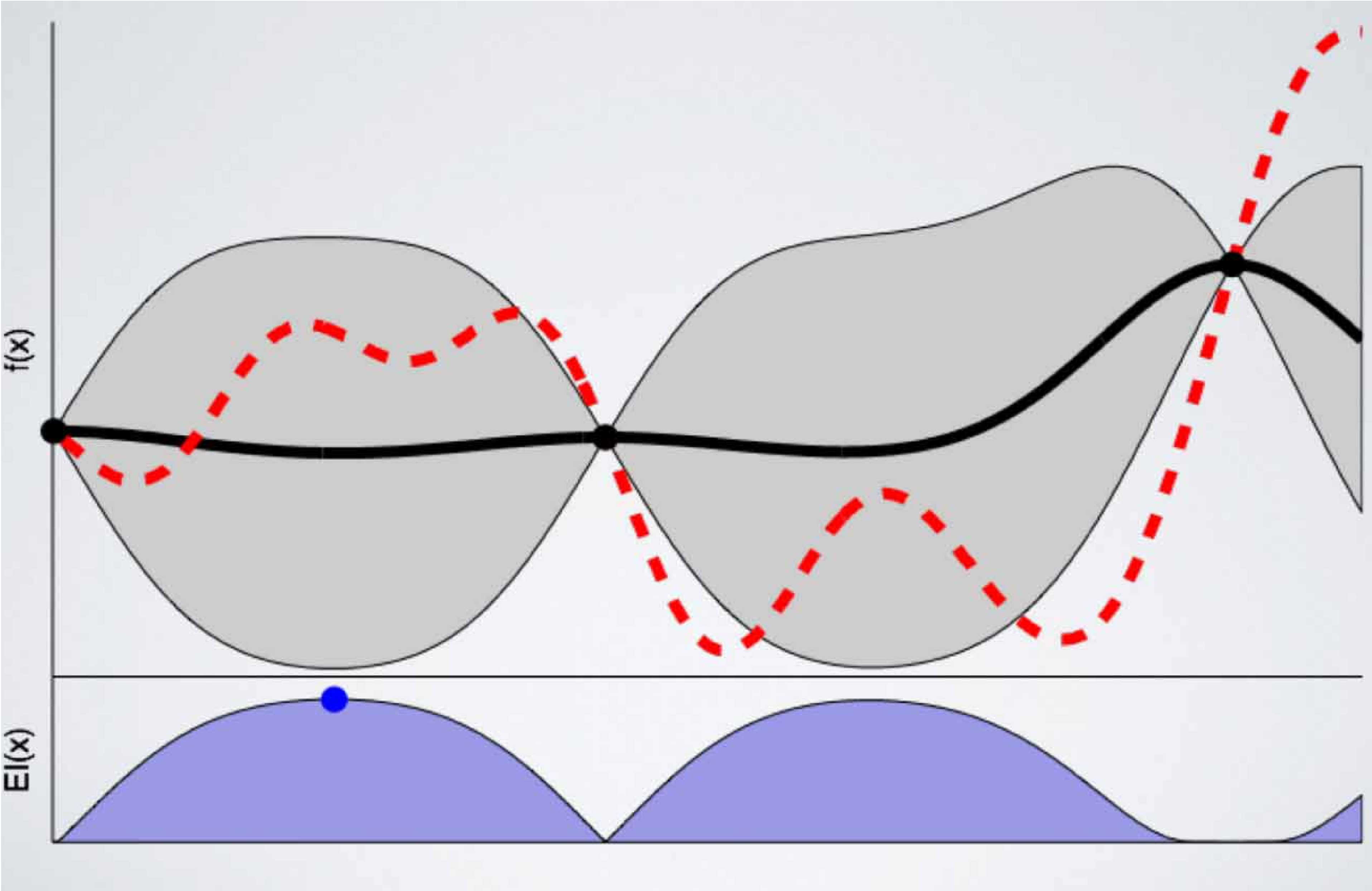


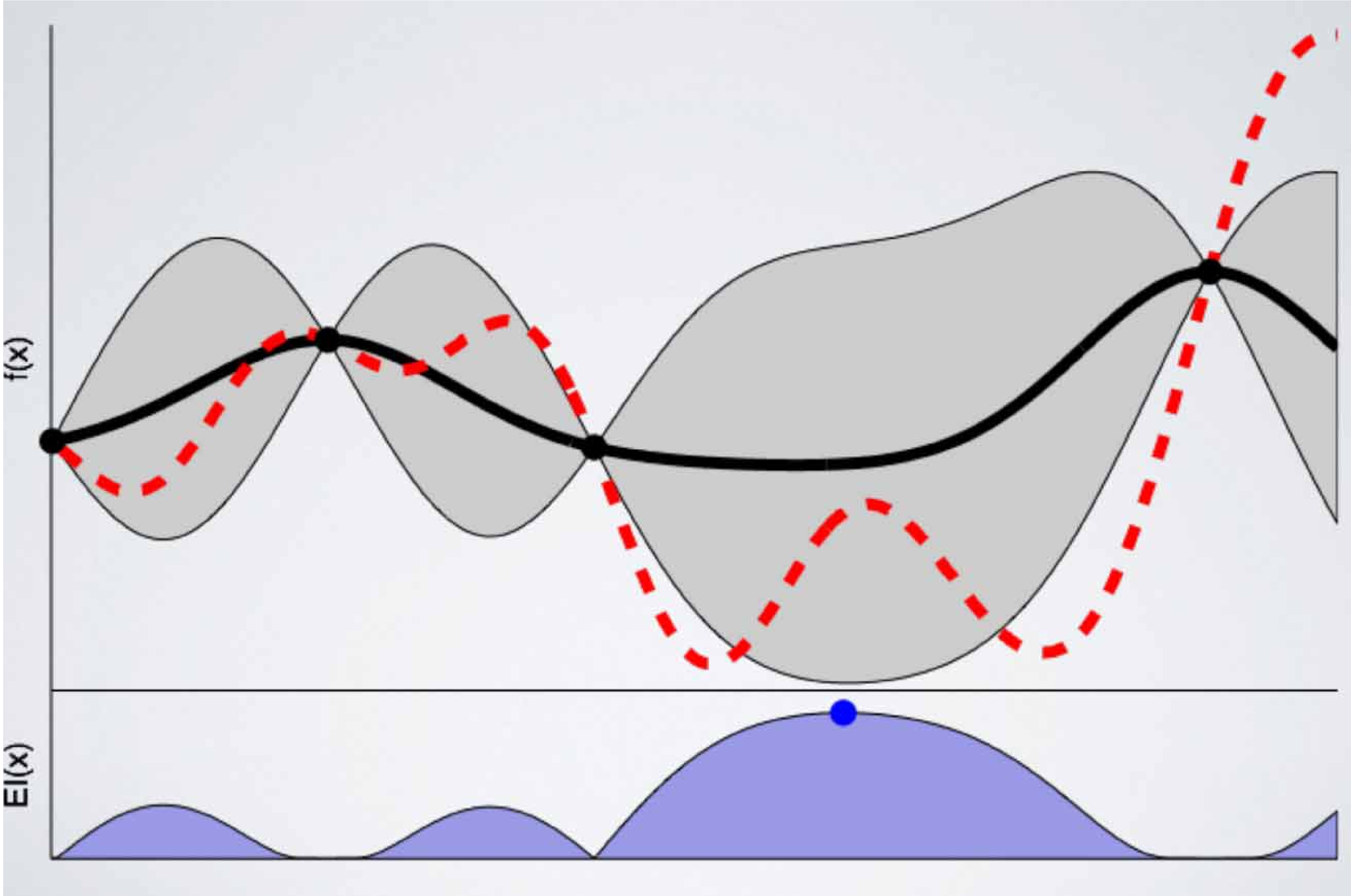
Brochu, E., Cora, V. M. & De Freitas, N. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv:1012.2599.

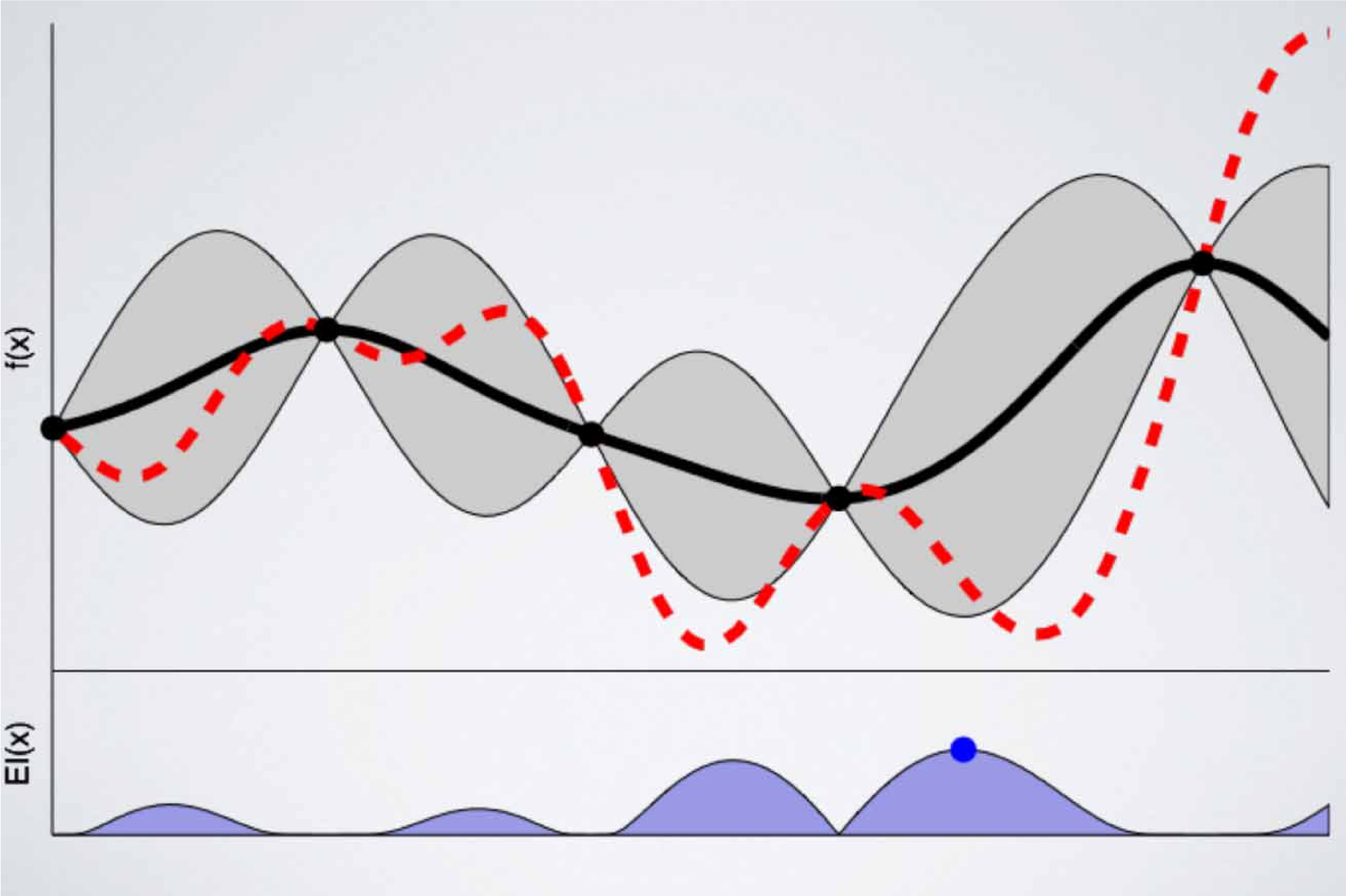
Demo on how Bayesian Optimization works …

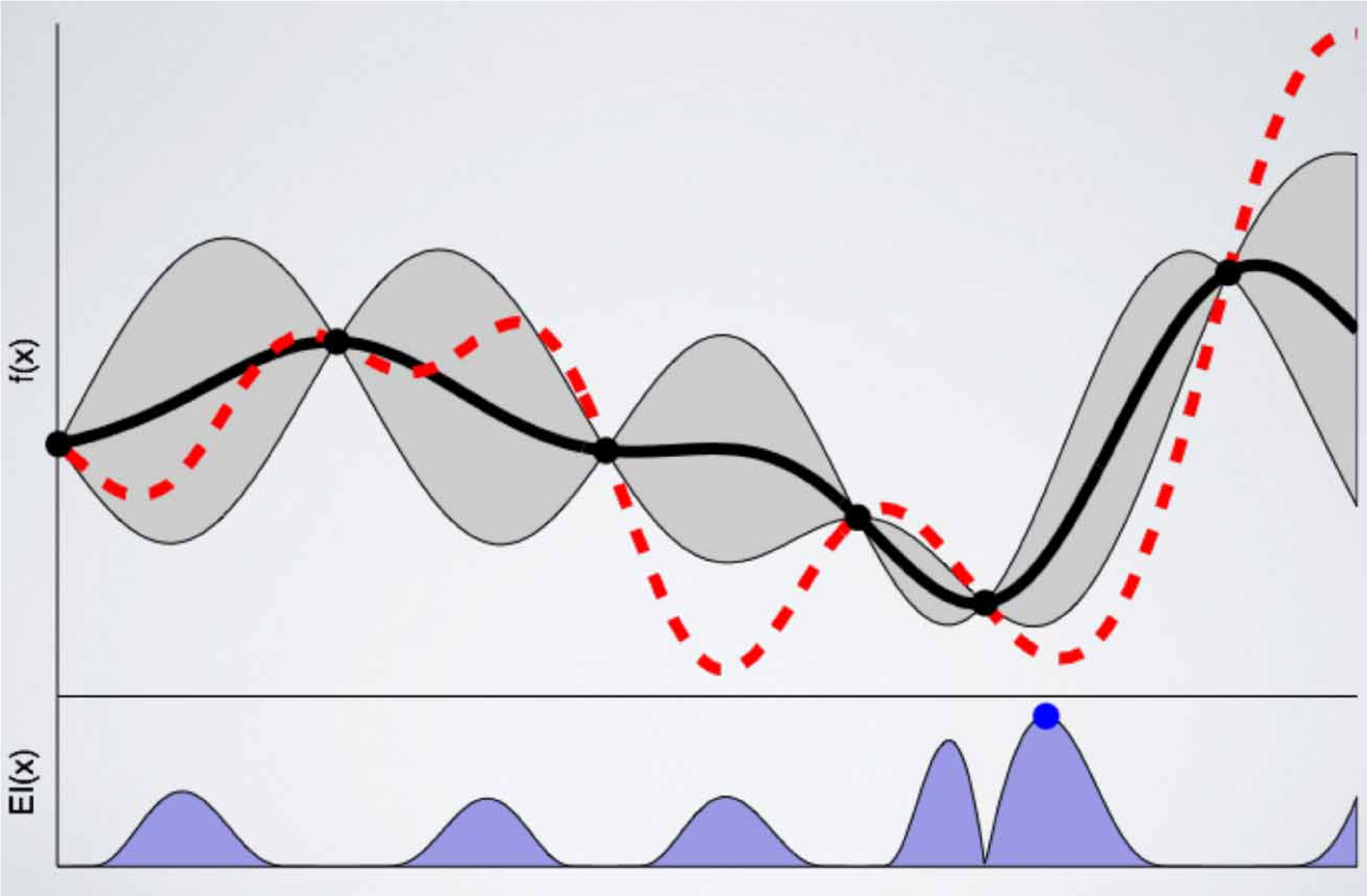


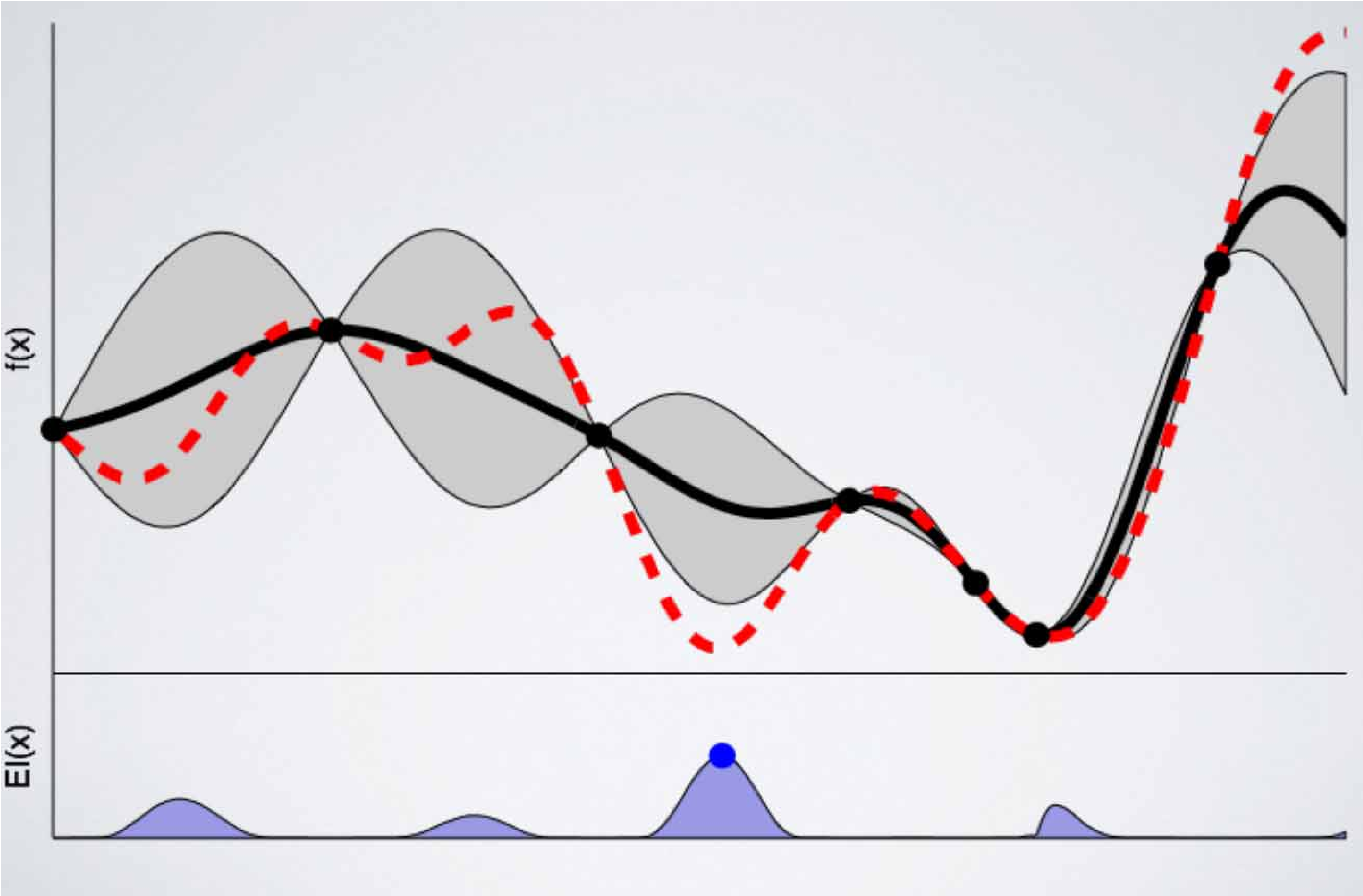
Snoek, J., Larochelle, H. & Adams, R. P. Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems, 2012. 2951-2959.

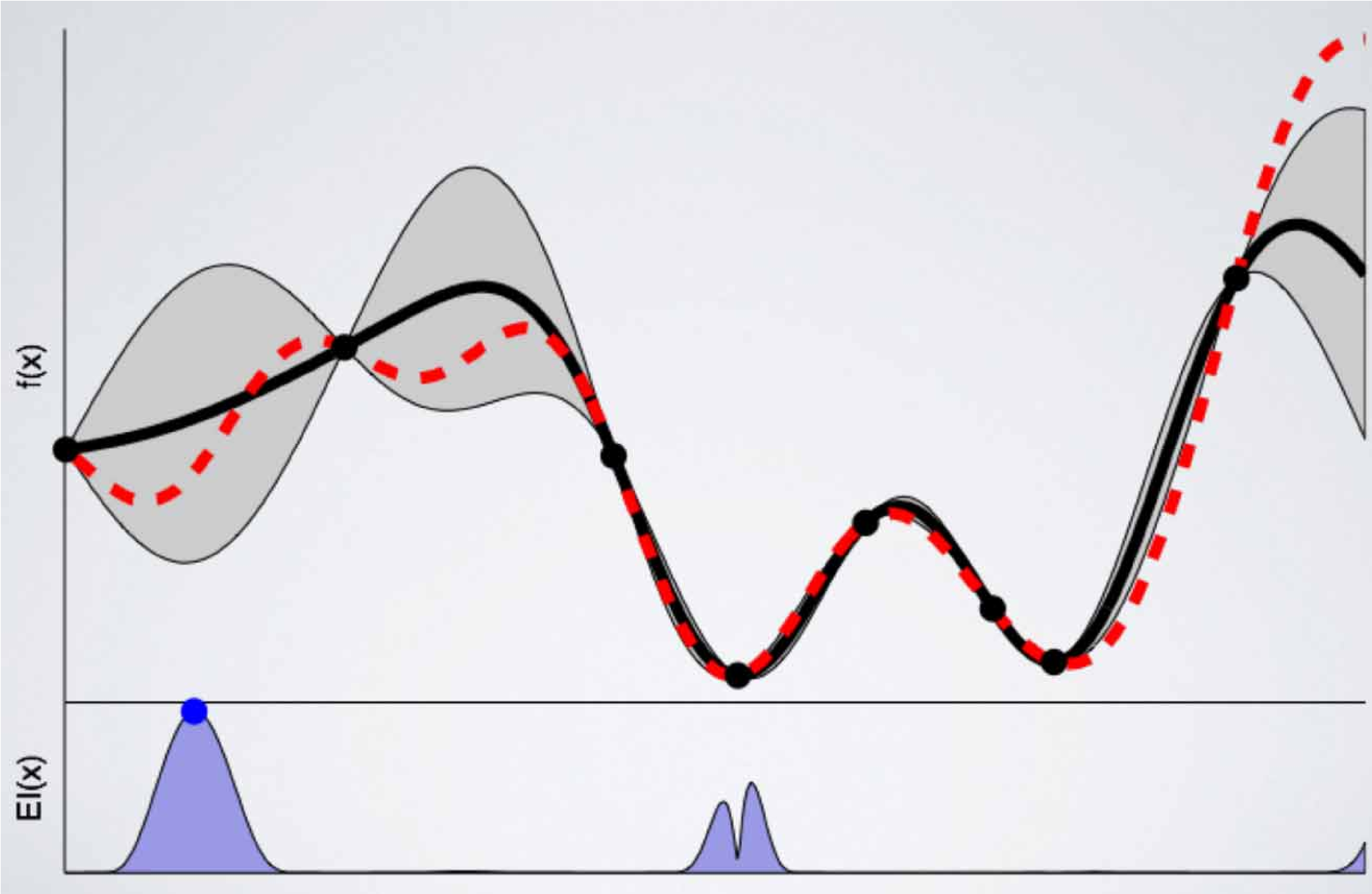

f(x)
EI(x)


f(x)
EI(x)


f(x)
EI(x)


f(x)
EI(x)


f(x)
EI(x)

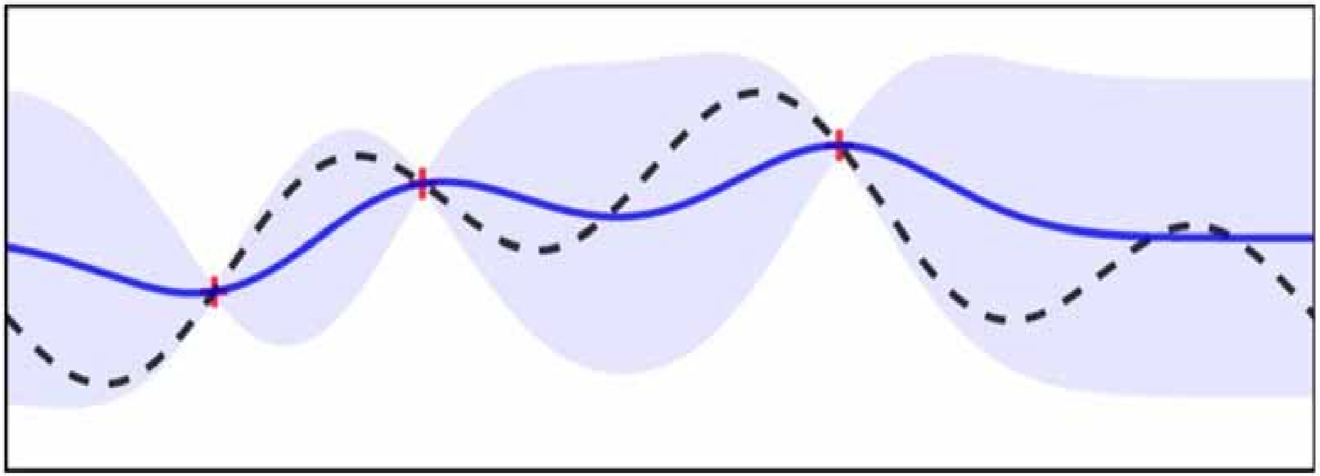

f(x)
EI(x)

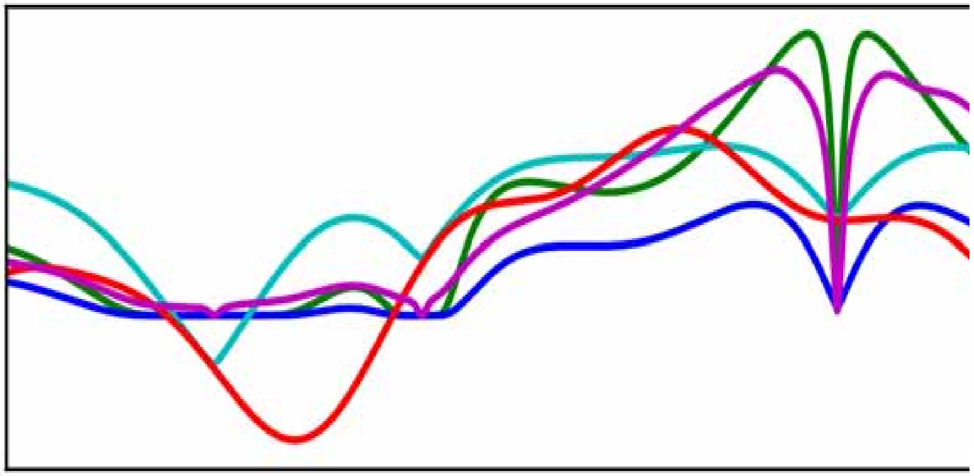
Why is this relevant for health informatics?

- Take patient information, e.g., observations, symptoms, test results, -omics data, etc. etc.
- Reach conclusions, and **predict** into the future, e.g. how likely will the patient be …
- Prior = belief before making a particular observation
- Posterior = belief after making the observation and is the prior for the next observation – intrinsically incremental

$$p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{\sum p(x_i, y_j)p(x_i)}$$

Grand Goal of automatic Machine Learning …





Algorithm 1: Bayesian optimization

1: **for** $n = 1, 2, \ldots,$ **do**
2: select new $\mathbf{x}_{n+1}$ by optimizing acquisition function α

$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}} \alpha(\mathbf{x}; \mathcal{D}_n)$$

3: query objective function to obtain y_{n+1}
4: augment data $\mathcal{D}_{n+1} = \{\mathcal{D}_n, (\mathbf{x}_{n+1}, y_{n+1})\}$
5: update statistical model
6: **end for**

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & De Freitas, N. 2016. **Taking the human out of the loop:** A review of Bayesian optimization. *Proceedings of the IEEE*, 104, (1), 148-175, doi:10.1109/JPROC.2015.2494218.

- Today most ML-applications are using automatic Machine Learning (aML) approaches

- automatic Machine Learning (aML) := algorithms which interact with agents and can optimize their learning behaviour trough this interaction

Best practice examples of aML

Wenger 600638 IBEX 17" Laptop Backpack with Tablet / eReader Pocket (Black / Blue)

von Wenger

EUR 66,99 Prime

Andere Angebote

EUR 60,00 neu (22 Angebote)

EUR 57,70 gebraucht (1 Angebot)

Nur Artikel von Wenger anzeigen

295

Für größere Ansicht Maus über das Bild ziehen

Lenovo Einsteiger Not
Arbeitsspeicher und W

von Lenovo

70 Kundenre

Bestseller Nr. 1 in Notebooks

Unverb. Preisempf.: ~~EUR 349,0(~~

Preis: EUR 273

Sie sparen: EUR 75,01

Alle Preisan

Lieferung Mittwoch, 6. Juli: Be

Details.

Auf Lager.

Verkauf und Versand durch Amaz

20 neu ab EUR 273,99 4 gebr

Größe: **500GB**

1TB **500GB**

Stil: **Intel Pentium**

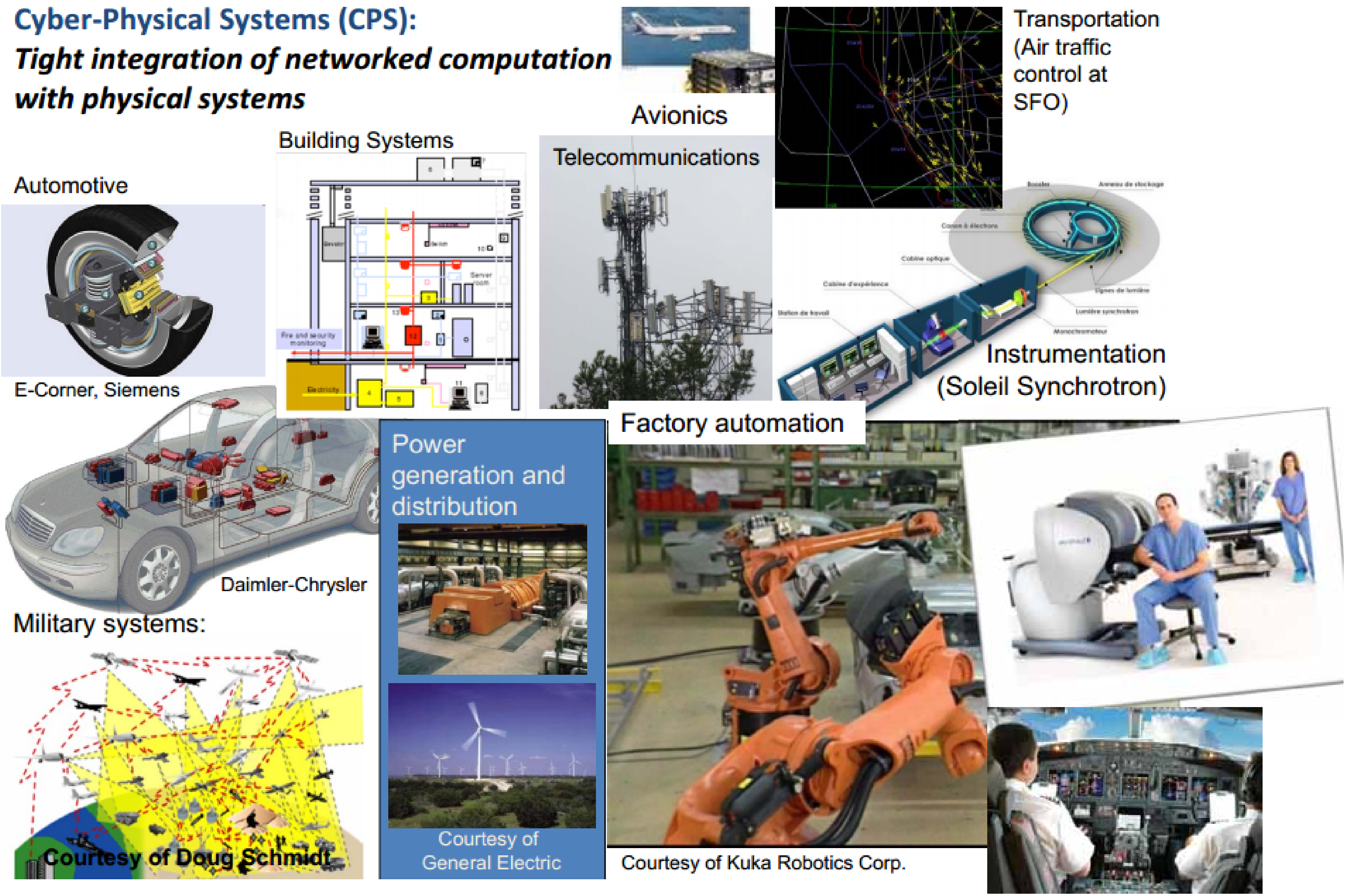
Intel Core i3 **Intel Pentiur**

- Prozessor: Intel Pentium N35
- Besonderheiten: HD Glare Di
- Akku: bis zu 4 Stunden Akkula
- Herstellergarantie: 12 Monate Angaben des jeweiligen Verkä
- Lieferumfang: Lenovo ideapac

› Weitere Produktdetails

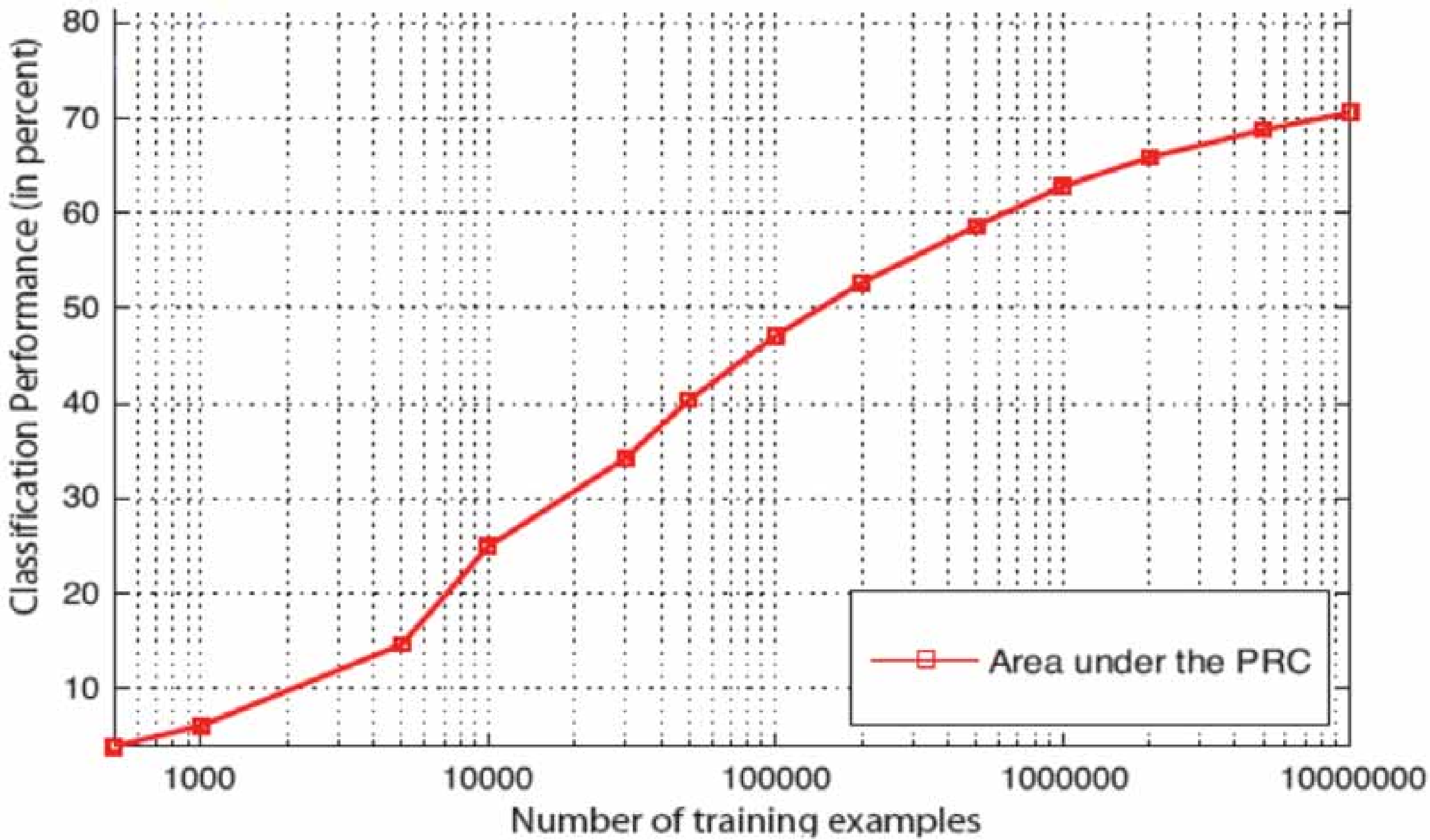
Fully automatic autonomous vehicles (Google car)

Dietterich, T. G. & Horvitz, E. J. 2015. Rise of concerns about AI: reflections and directions. Communications of the ACM, 58, (10), 38-40.



Seshia, S. A., Juniwal, G., Sadigh, D., Donze, A., Li, W., Jensen, J. C., Jin, X., Deshmukh, J., Lee, E. & Sastry, S. 2015. Verification by, for, and of Humans: Formal Methods for Cyber-Physical Systems and Beyond. Illinois ECE Colloquium.

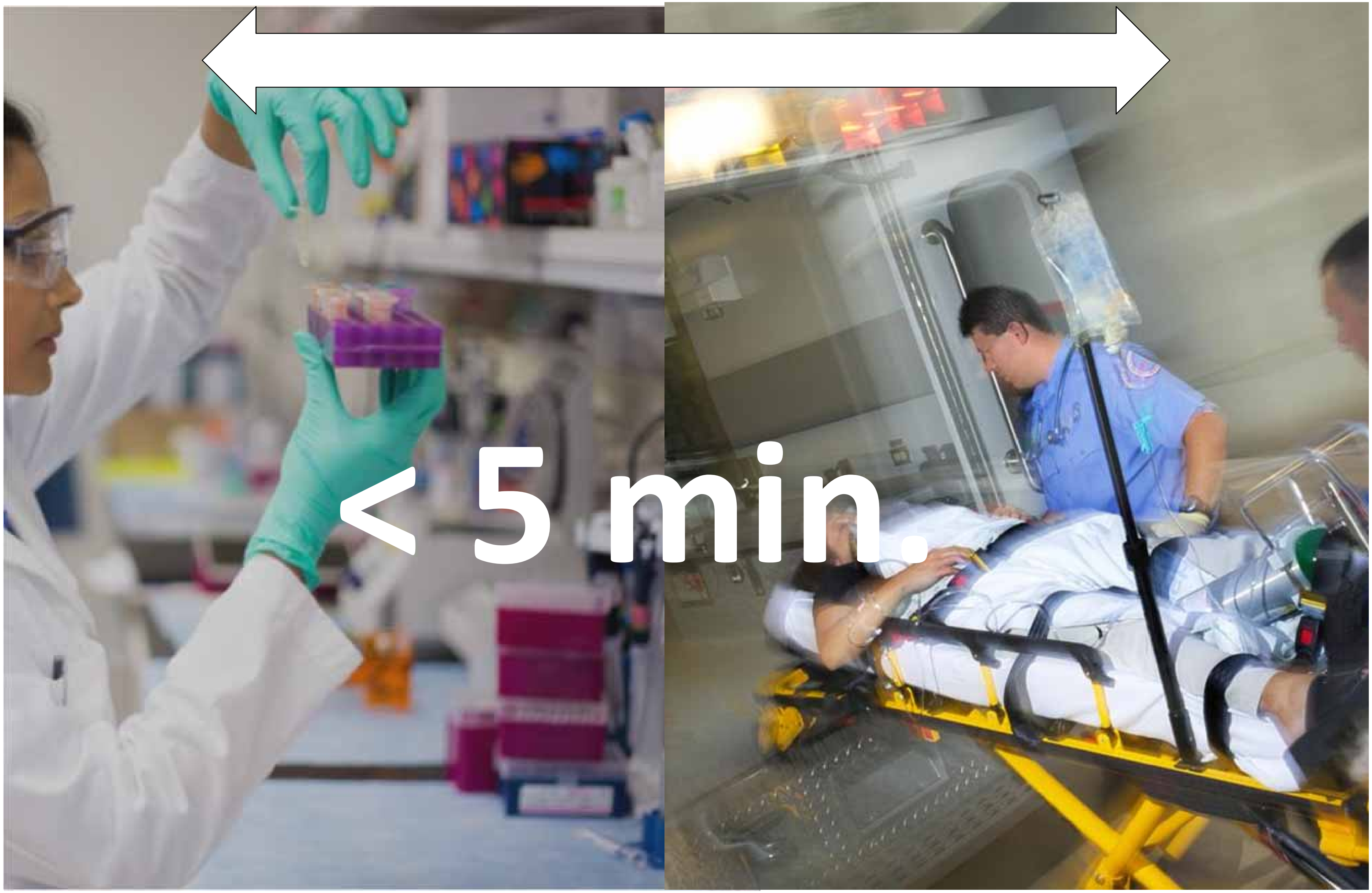
Big Data is necessary for aML !



Sonnenburg, S., Rätsch, G., Schäfer, C. & Schölkopf, B. 2006. Large scale multiple kernel learning. Journal of Machine Learning Research, 7, (7), 1531-1565.


Does this work here
as well?

Medical Decision Making as a Search Task in $\mathcal{H}$ Problem: Time (t)


< 5 min.

- Sometimes we **do not have "big data",** where aML-algorithms benefit.
- Sometimes we have
  - **Small amount of data sets**
  - Rare Events – **no training samples**
  - **NP-hard problems, e.g.**
    - Subspace Clustering,
    - Protein-Folding,
    - k-Anonymization,
    - Graph Coloring, Category Discovery, etc. etc….

Hans Holbein d.J., 1533,
The Ambassadors,
London: National Gallery

Lopez-Paz, D., Muandet, K., Schölkopf, B. & Tolstikhin, I. 2015. Towards a learning theory of cause-effect inference. Proceedings of the 32nd International Conference on Machine Learning, JMLR, Lille, France.

https://www.youtube.com/watch?v=9KiVNIUMmCc

Sometimes we (still) need a human-in-the-loop

A doctor-in-the-loop

- interactive Machine Learning (iML) := algorithms which interact with agents*) and can optimize their learning behaviour trough this interaction
- *)**where the agents can be human**

Holzinger, A. 2015. Interactive Machine Learning (iML). Informatik Spektrum
DOI: 10.1007/s00287-015-0941-6

A group of experts-in-the-loop

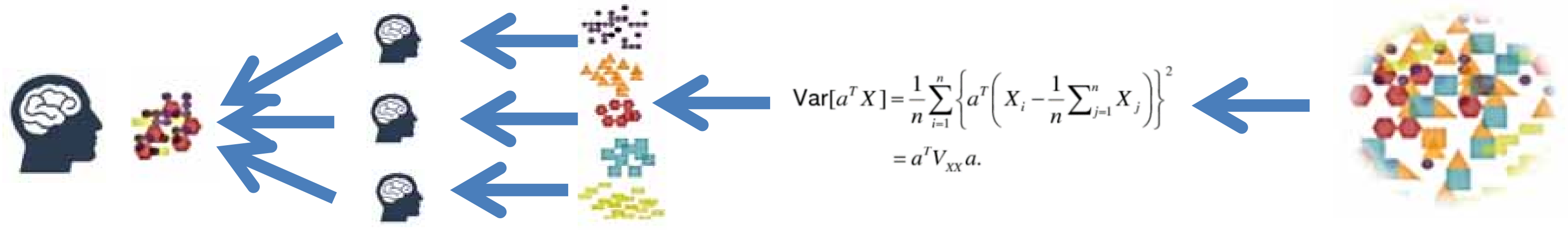
A crowd of people-in-the-loop

Multi-Agents

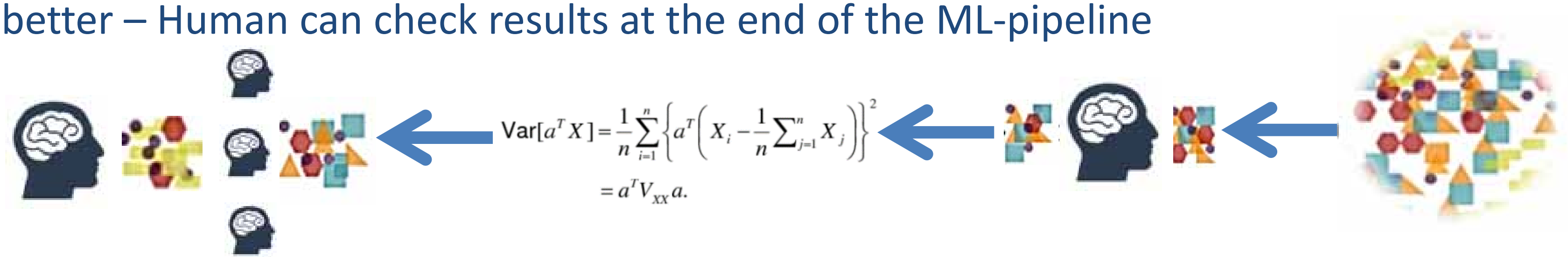
Nature-inspired agents

aML: Unsupervised – Supervised – Semi-supervised

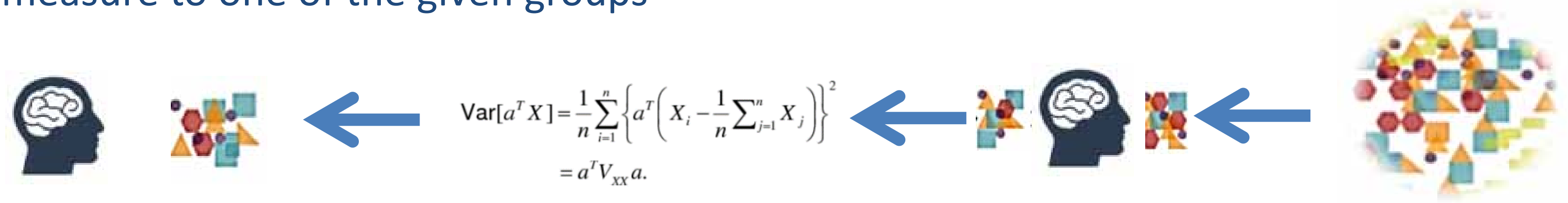
A) Unsupervised ML: Algorithm is applied on the raw data and learns fully automatic – Human can check results at the end of the ML-pipeline



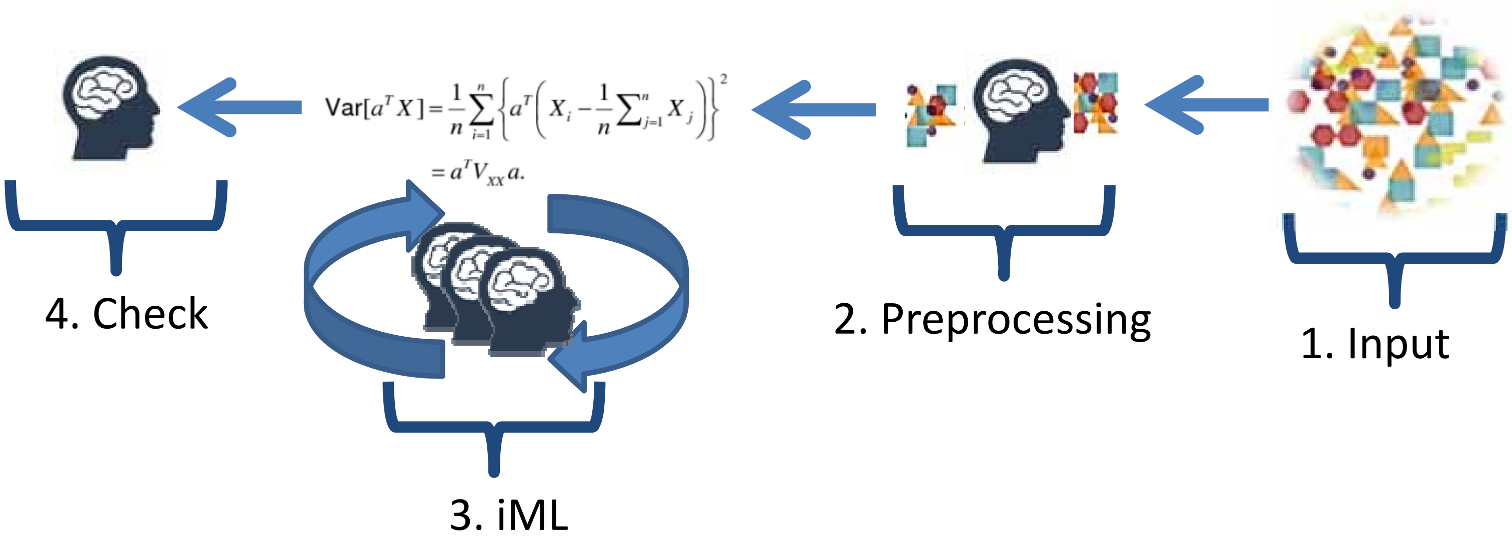
B) Supervised ML: Humans are providing the labels for the training data and/or select features to feed the algorithm to learn – the more samples the better – Human can check results at the end of the ML-pipeline



C) Semi-Supervised Machine Learning: A mixture of A and B – mixing labeled and unlabeled data so that the algorithm can find labels according to a similarity measure to one of the given groups



D) **Interactive Machine Learning:** Human is seen as an agent involved in the actual learning phase, step-by-step influencing measures such as distance, cost functions …



Constraints of humans: Robustness, subjectivity, transfer?

Open Questions: Evaluation, replicability, …

Holzinger, A. 2016. Interactive Machine Learning for Health Informatics: When do we need the human-in-the-loop? Brain Informatics (BRIN), 3, (2), 119-131, doi:10.1007/s40708-016-0042-6.

- Example 1: k-Anonymity
- Example 2: Protein Folding
- Example 3: Subspace Clustering

A3

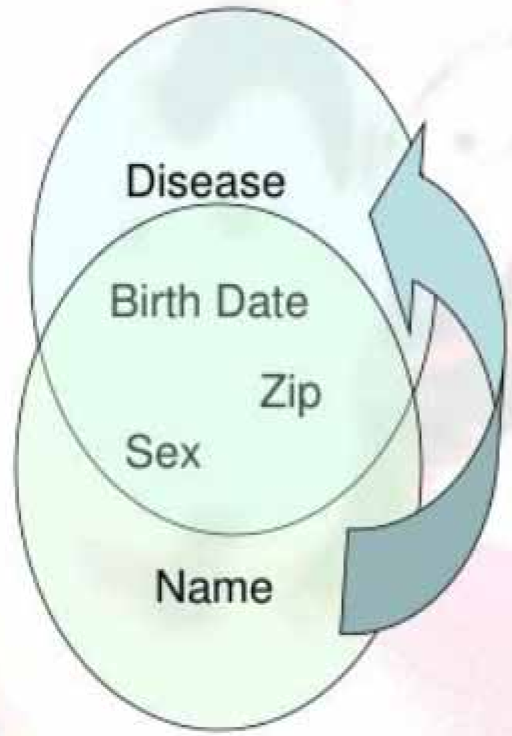
87 % of the population in the USA can be uniquely re-identified by Zip-Code, Gender and date of birth

Hospital Patient Data

Birthdate	Sex	Zipcode	Disease
1/21/76	Male	53715	Flu
4/13/86	Female	53715	Hepatitis
2/28/76	Male	53703	Brochitis
1/21/76	Male	53703	Broken Arm
4/13/86	Female	53706	Sprained Ankle
2/28/76	Female	53706	Hang Nail

Voter Registration Data

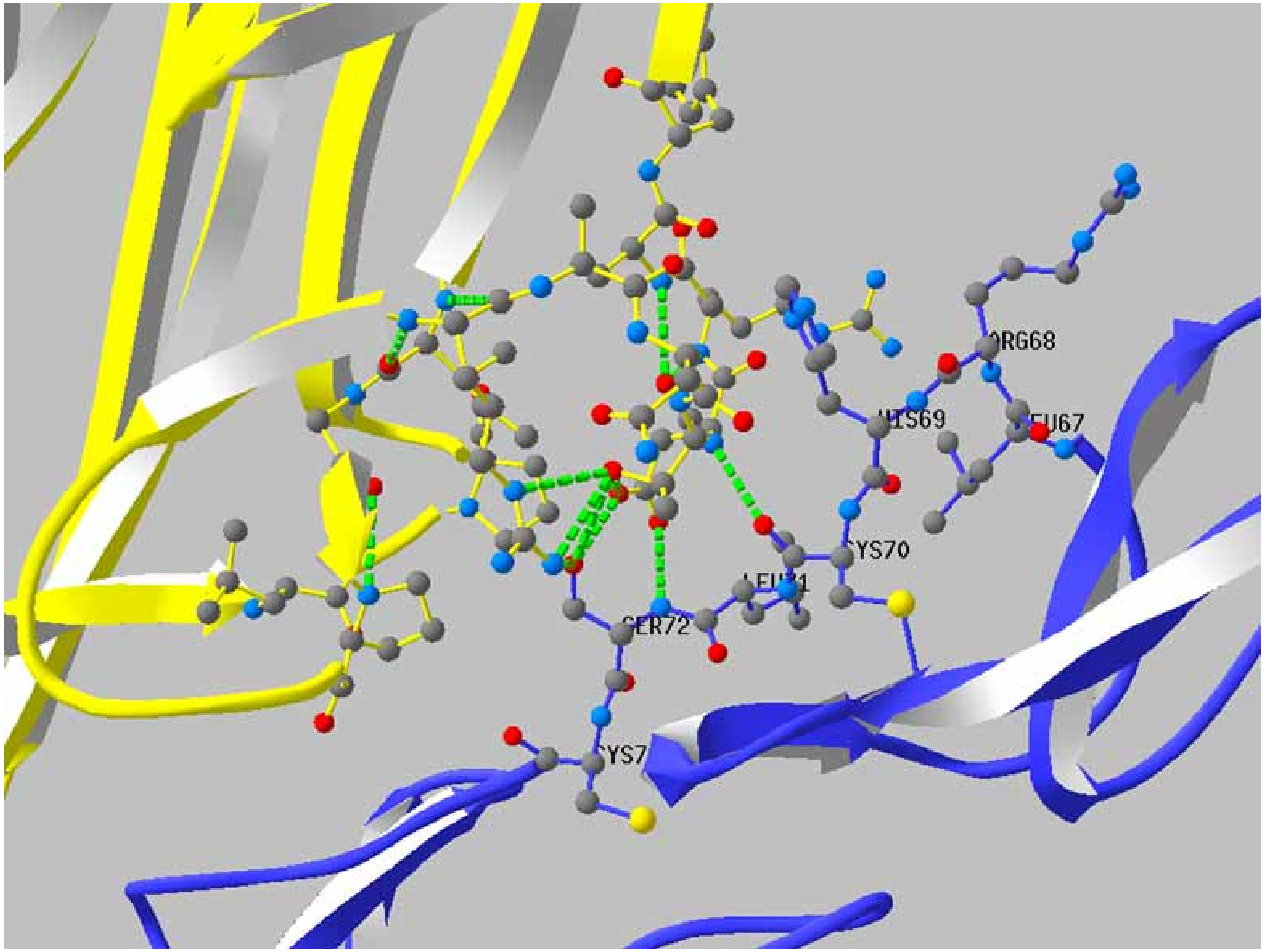
Name	Birthdate	Sex	Zipcode
Andre	1/21/76	Male	53715
Beth	1/10/81	Female	55410
Carol	10/1/44	Female	90210
Dan	2/21/84	Male	02174
Ellen	4/19/72	Female	02237



Sweeney, L. 2002. Achieving k-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10, (05), 571-588.

Proteins are the building blocks of life …

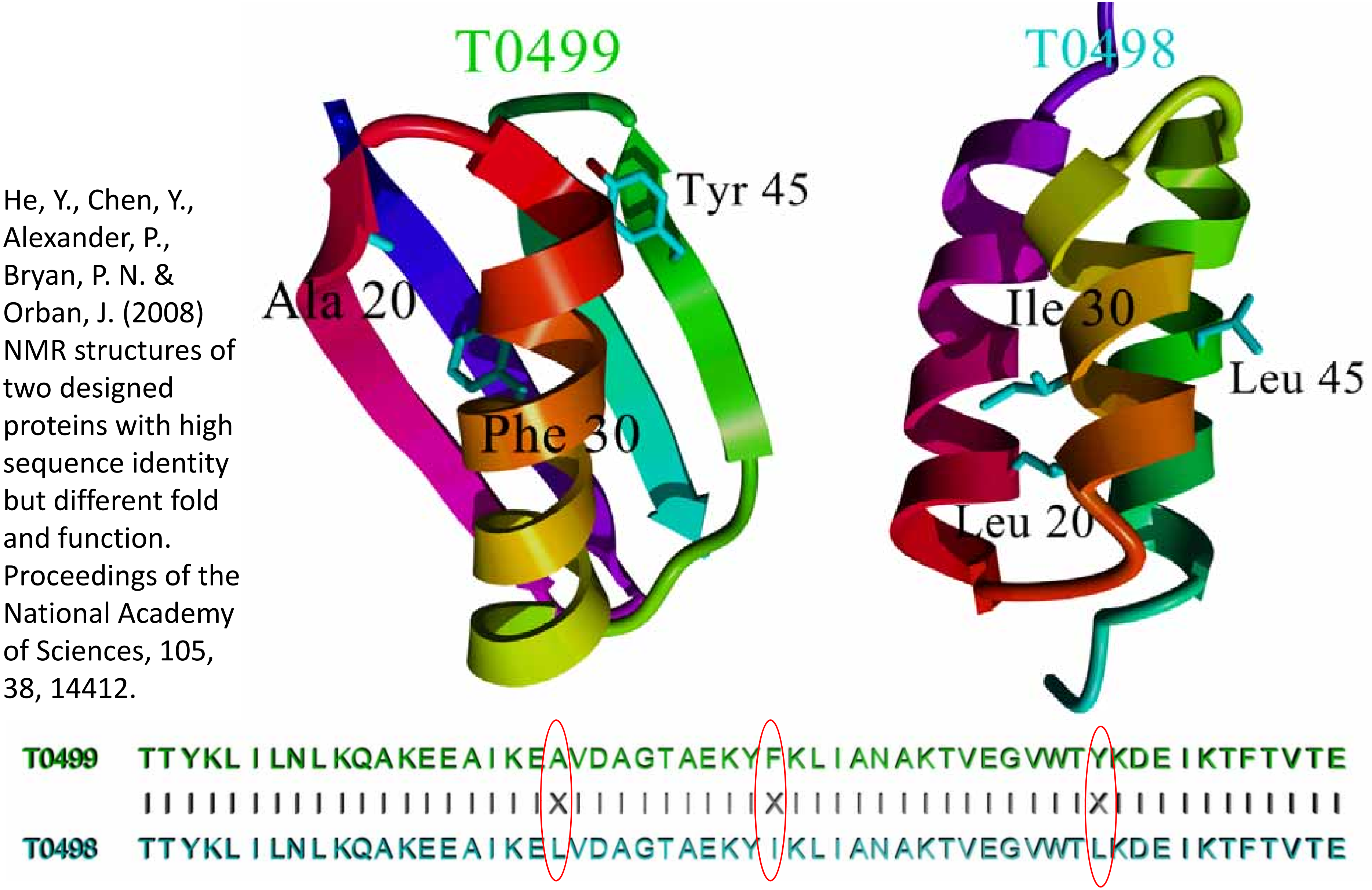
A4

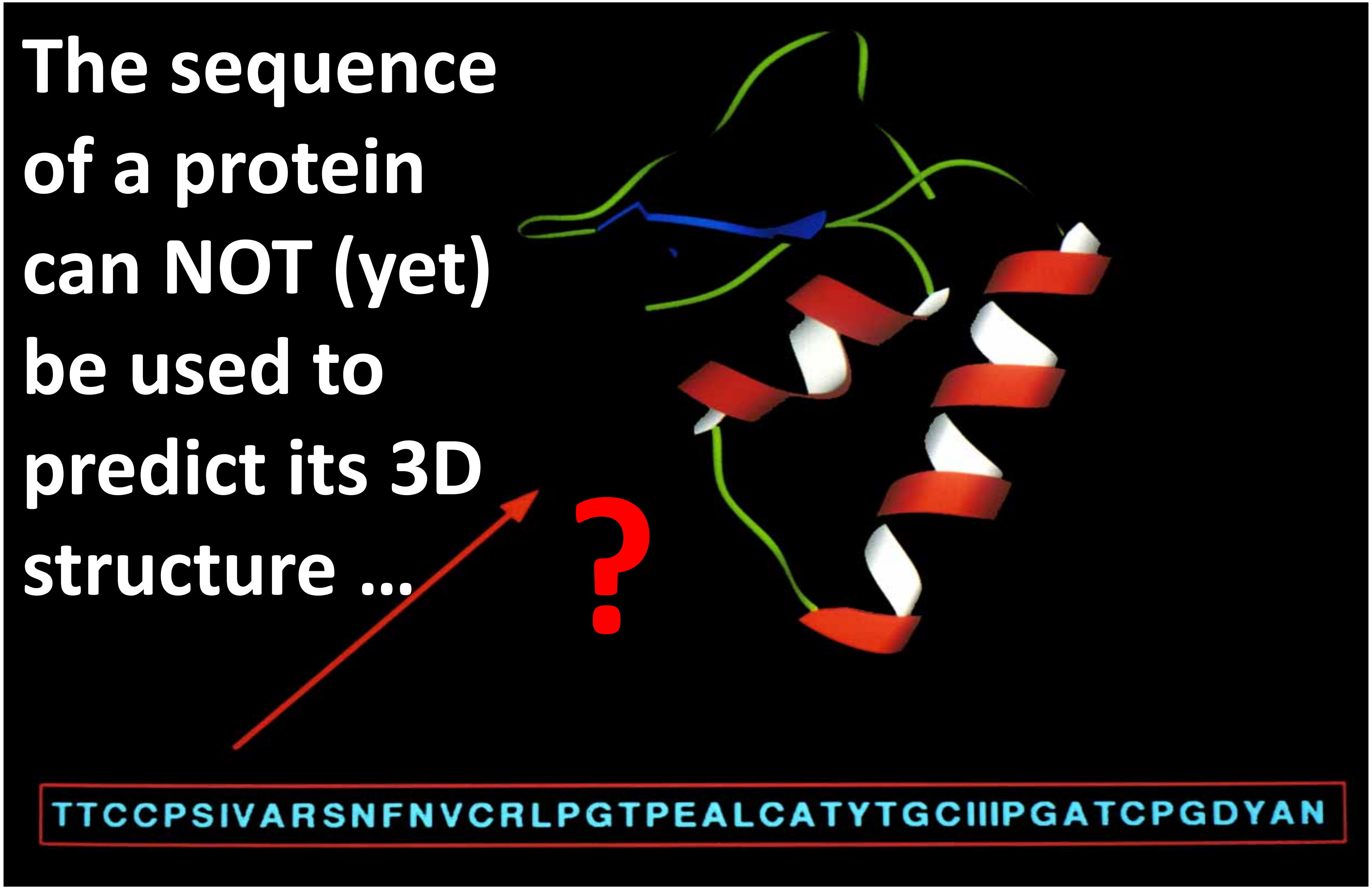


Wiltgen, M., Holzinger, A. & Tilz, G. P. (2007) Interactive Analysis and Visualization of Macromolecular Interfaces Between Proteins. In: *Lecture Notes in Computer Science (LNCS 4799). Berlin, Heidelberg, New York, Springer, 199-212.*

Example 2 Protein Folding: ὁμολογέω (homologeo)

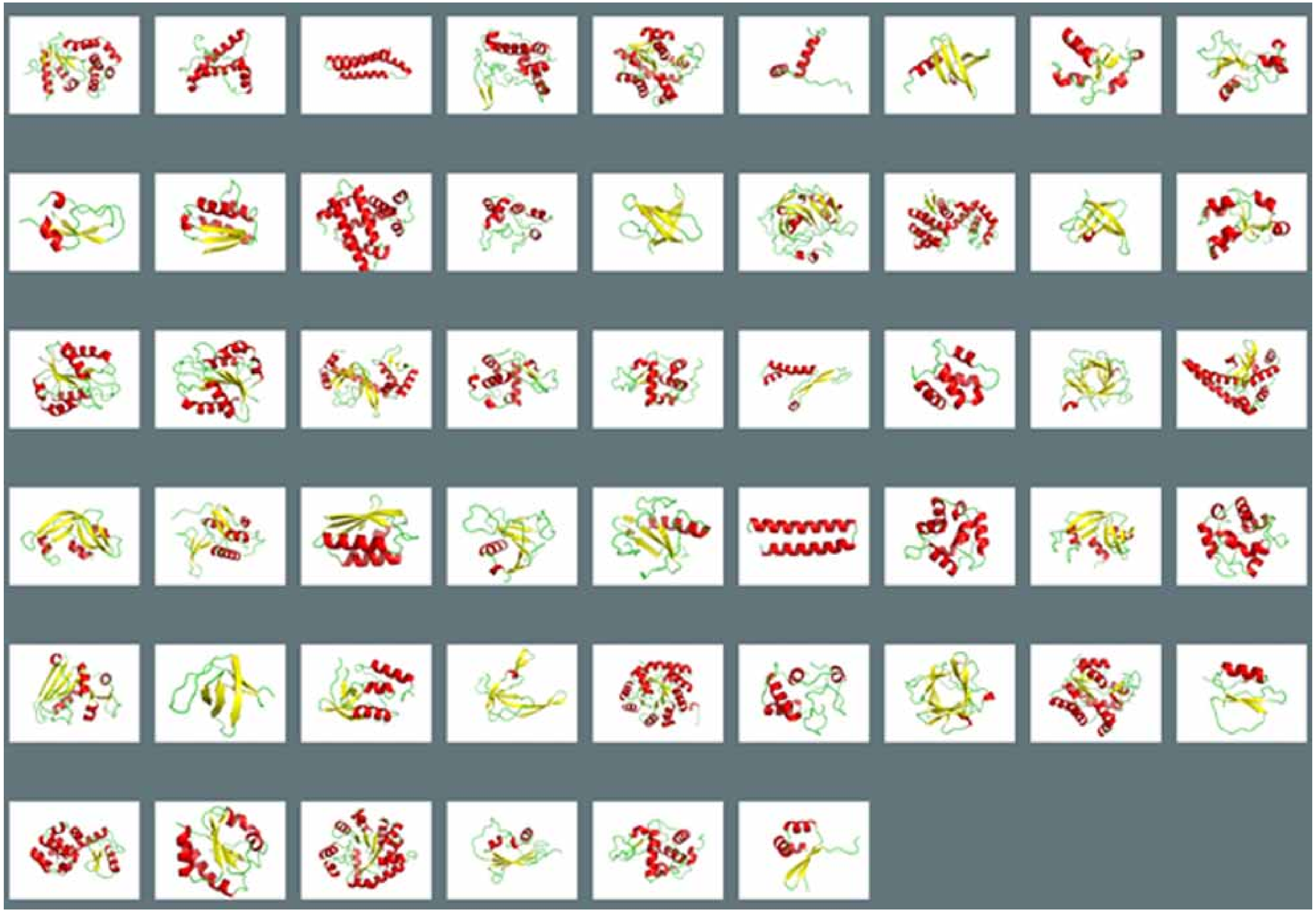
He, Y., Chen, Y., Alexander, P., Bryan, P. N. & Orban, J. (2008) NMR structures of two designed proteins with high sequence identity but different fold and function. Proceedings of the National Academy of Sciences, 105, 38, 14412.





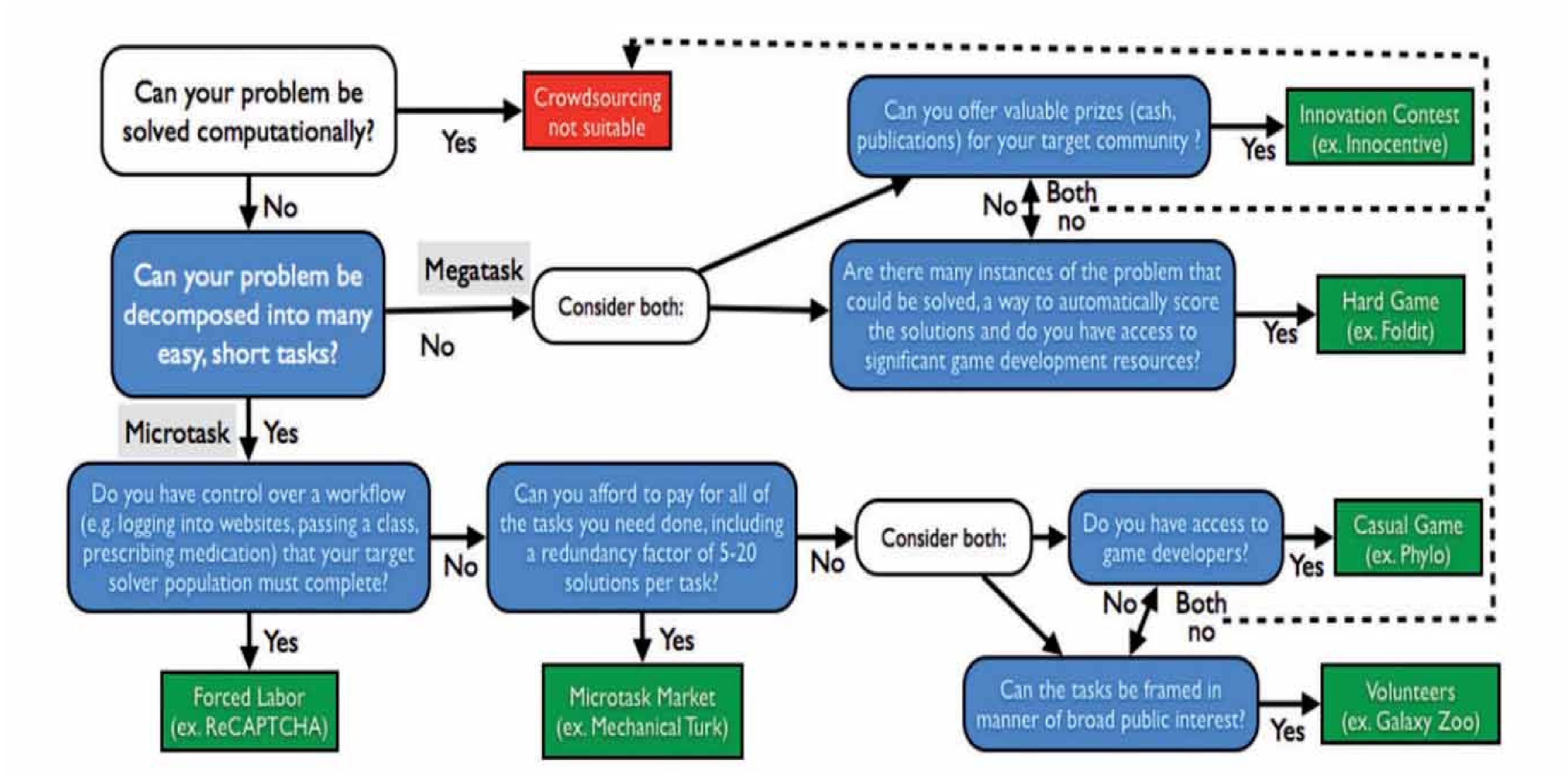
Anfinsen, C. B. **1973.** Principles that Govern the Folding of Protein Chains. Science, 181, (4096), 223-230.

State-of-2015: aML still is not able to help sufficiently …



Jia, L., Yarlagadda, R. & Reed, C. C. 2015. Structure Based Thermostability Prediction Models for Protein Single Point Mutations with Machine Learning Tools. Plos One, 10, (9).

Humans can help here: Crowdsourcing

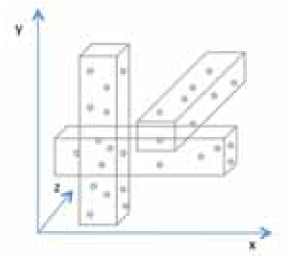


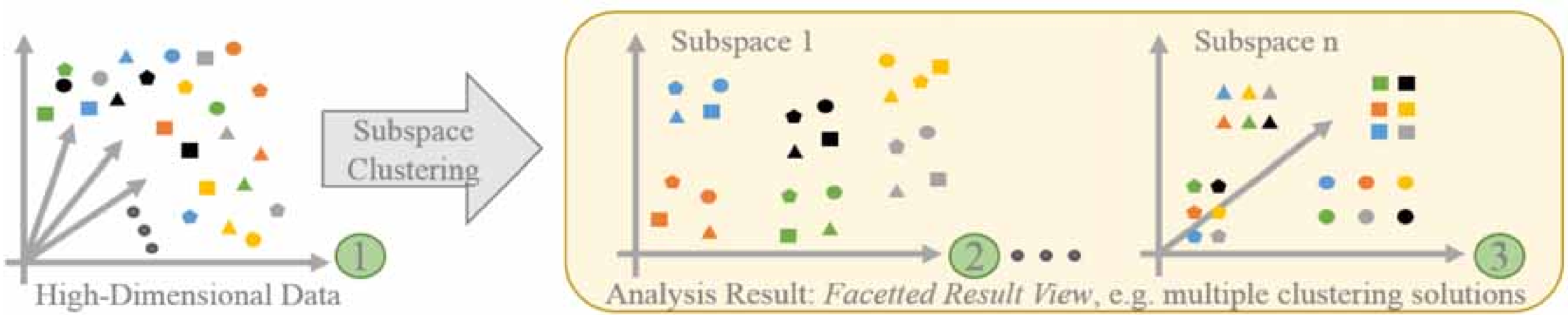
Good, B. M. & Su, A. I. 2013. Crowdsourcing for bioinformatics. *Bioinformatics,* 29, (16), 1925-1933.

Example 3: Subspace Clustering

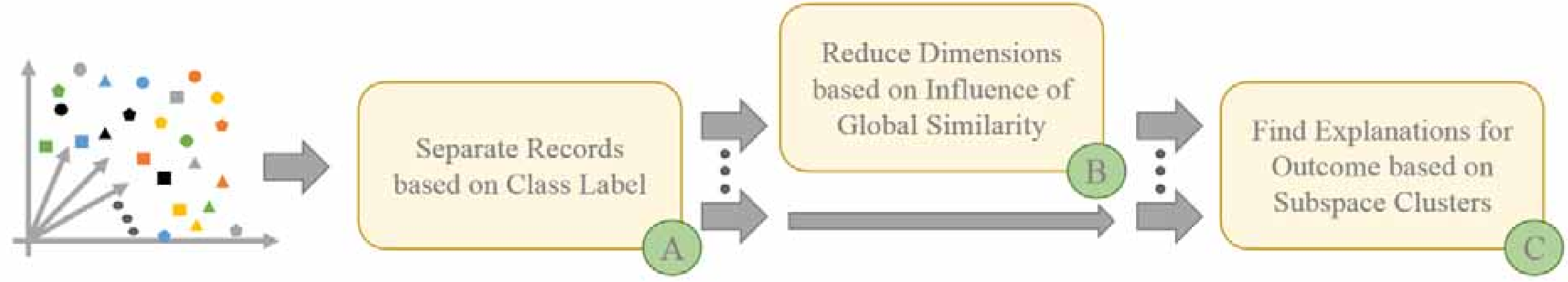
A5

- **Patterns may be found in subspaces (dimension combinations)**
- **Clustering and subset selection: Non-convex & NP-hard**
- **Real data are often noisy and corrupted**
- **Little prior knowledge about low-dim structures**
- **Data points in different subgroups can be very close**





Hund, M., Sturm, W., Schreck, T., Ullrich, T., Keim, D., Majnaric, L. & Holzinger, A. 2015. Analysis of Patient Groups and Immunization Results Based on Subspace Clustering. In: Lecture Notes in Artificial Intelligence LNAI 9250, pp. 358-368.



- **Positive** subspace clusters
  - One homogeneous cluster (healthy patients)
    - `hyper, CVD, neoplasm, psy.disorder, drug allergy`
    - No medications: `statins, anticoagulants, analgesics` and `clear` (preserved renal function)
- **Negative** subspace clusters
  - Cluster with obvious reasons for neg. outcome
    - Impairment of certain pathophysiologic mechanism increased `MCV`, decreased `VITB12, FOLNA, CORTIS`) despite no: `DM, drug allergy, Fglu, E/HB` (anaemia)

Human kernel …

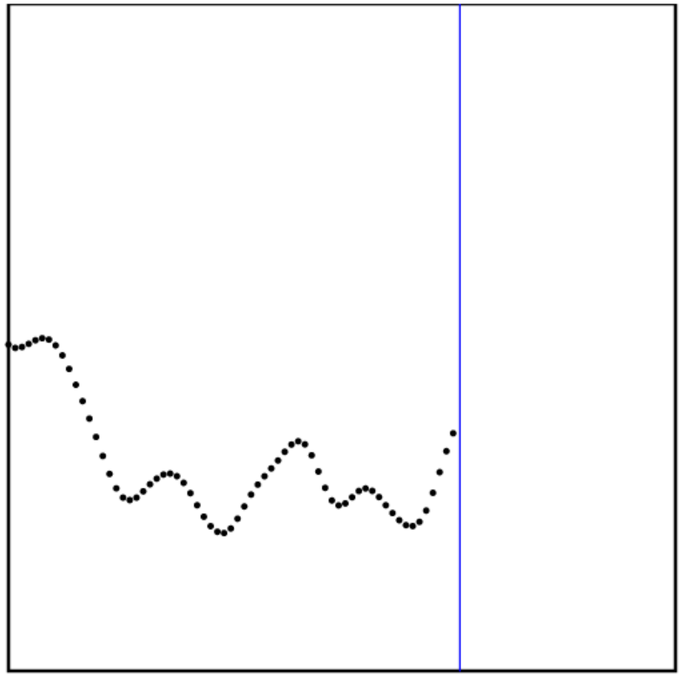
A2

Judgment 1 out of 33

This is the first function from the system. Please try to predict the new points as well as you can based on the points you can see.

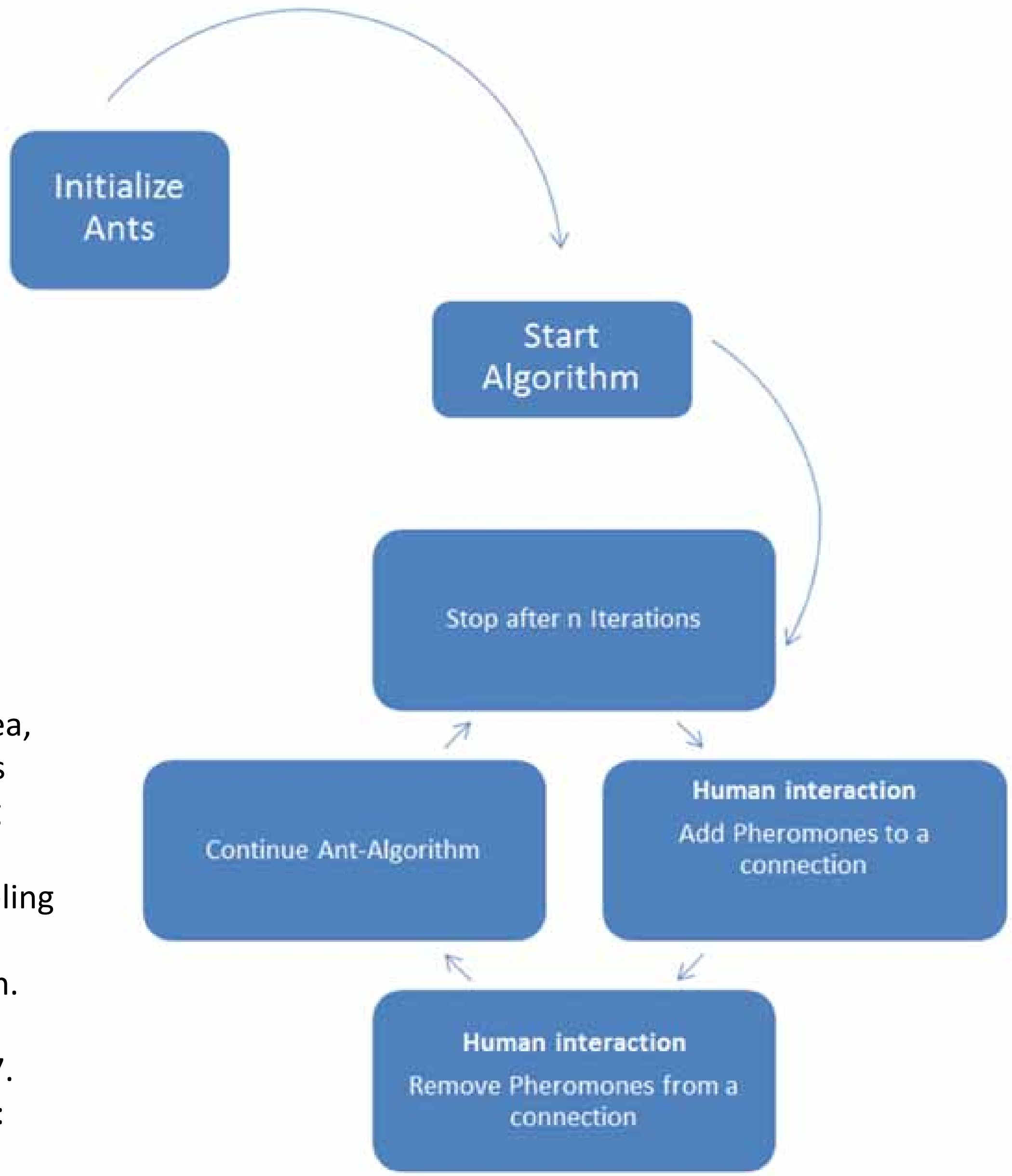
Please click along the blue line to say what you think the height of the point is for that location.

Once you have selected a position along the line, **hit the 's' key to submit the point**.



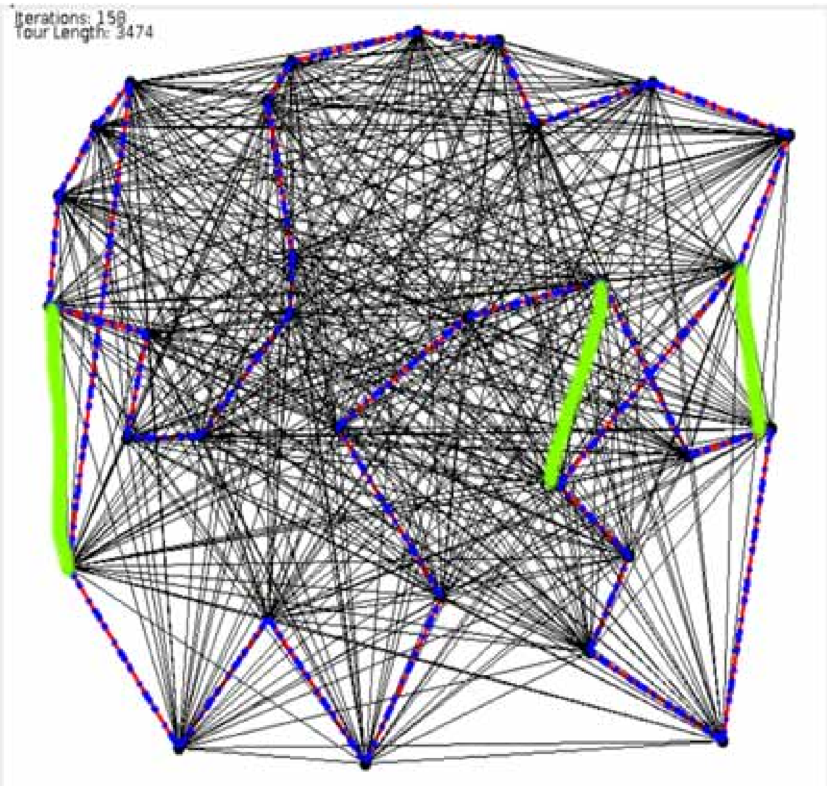
Wilson, A. G., Dann, C., Lucas, C. & Xing, E. P. The Human Kernel. Advances in Neural Information Processing Systems, 2015. 2836-2844.

$$\hat{p}^{a}_{bc} = \frac{\mu + \delta_{ac}}{2\mu + \delta_{ab} + \delta_{ac}} \quad \text{and} \quad K_{ii} = 1,$$



Holzinger, A., Plass, M., Holzinger, K., Crisan, G., Pintea, C. & Palade, V. 2016. Towards interactive Machine Learning (iML): Applying Ant Colony Algorithms to solve the Traveling Salesman Problem with the Human-in-the-Loop approach. Springer Lecture Notes in Computer Science LNCS 9817. Heidelberg, Berlin, New York: Springer, pp. in print.

A1



Algorithm 2: Ant Colony Algorithm iML

Input : ProblemSize, $Population_{size}$, m, ρ, β, σ, $q0$
Output: P_{best}
$P_{best} \leftarrow$ CreateHeuristicSolution(ProblemSize);
$Pbest_{cost} \leftarrow$ Cost(S_h);
$Pheromone_{init} \leftarrow \frac{1.0}{ProblemSize \times Pbest_{const}}$;
$Pheromone \leftarrow$ InitializePheromone($Pheromone_{init}$);
while $\neg$*StopCondition()* **do**
  for $i = 1$ to m **do**
    $S_i \leftarrow$ ConstructSolution(Pheromone, ProblemSize, β, $q0$);
    $Si_{cost} \leftarrow$ Cost(S_i);
    if $Si_{cost} \leq$ Pbest_{cost} **then**
      $Pbest_{cost} \leftarrow Si_{cost}$;
      $P_{best} \leftarrow S_i$;
    end
    LocalUpdateAndDecayPheromone(Pheromone, S_i, Si_{cost}, σ);
  end
  GlobalUpdateAndDecayPheromone(Pheromone, P_{best}, $Pbest_{cost}$, ρ);
  while *isUserInteraction()* **do**
    GlobalAddAndRemovePheromone(Pheromone, P_{best}, $Pbest_{cost}$, ρ);
  end
end
return P_{best};

http://hci-kdd.org/project/iml/

Holzinger, A., Plass, M., Holzinger, K., Crisan, G., Pintea, C. & Palade, V. 2016. Towards interactive Machine Learning (iML): Applying Ant Colony Algorithms to solve the Traveling Salesman Problem with the Human-in-the-Loop approach. Springer Lecture Notes in Computer Science LNCS 9817. Heidelberg, Berlin, New York: Springer, pp. in print.

- ❶ Heterogeneous data sources
  - need for data integration and data fusion
- ❷ Complexity – reduction of search space
  - combining the best of Human & Computer
- ❸ What is interesting? – and relevant!
  - need of effective mapping $\mathbb{R}^N \rightarrow \mathbb{R}^2$
- ❹ Clinical time limits “5 Minutes”
  - need of efficient solutions

Holzinger, A. & Jurisica, I. 2014. Knowledge Discovery and Data Mining in Biomedical Informatics: The future is in Integrative, Interactive Machine Learning Solutions In: LNCS 8401. Heidelberg, Berlin: Springer, pp. 1-18.

Multi-Task Learning (MTL)

for improving prediction performance, help to reduce **catastrophic forgetting**

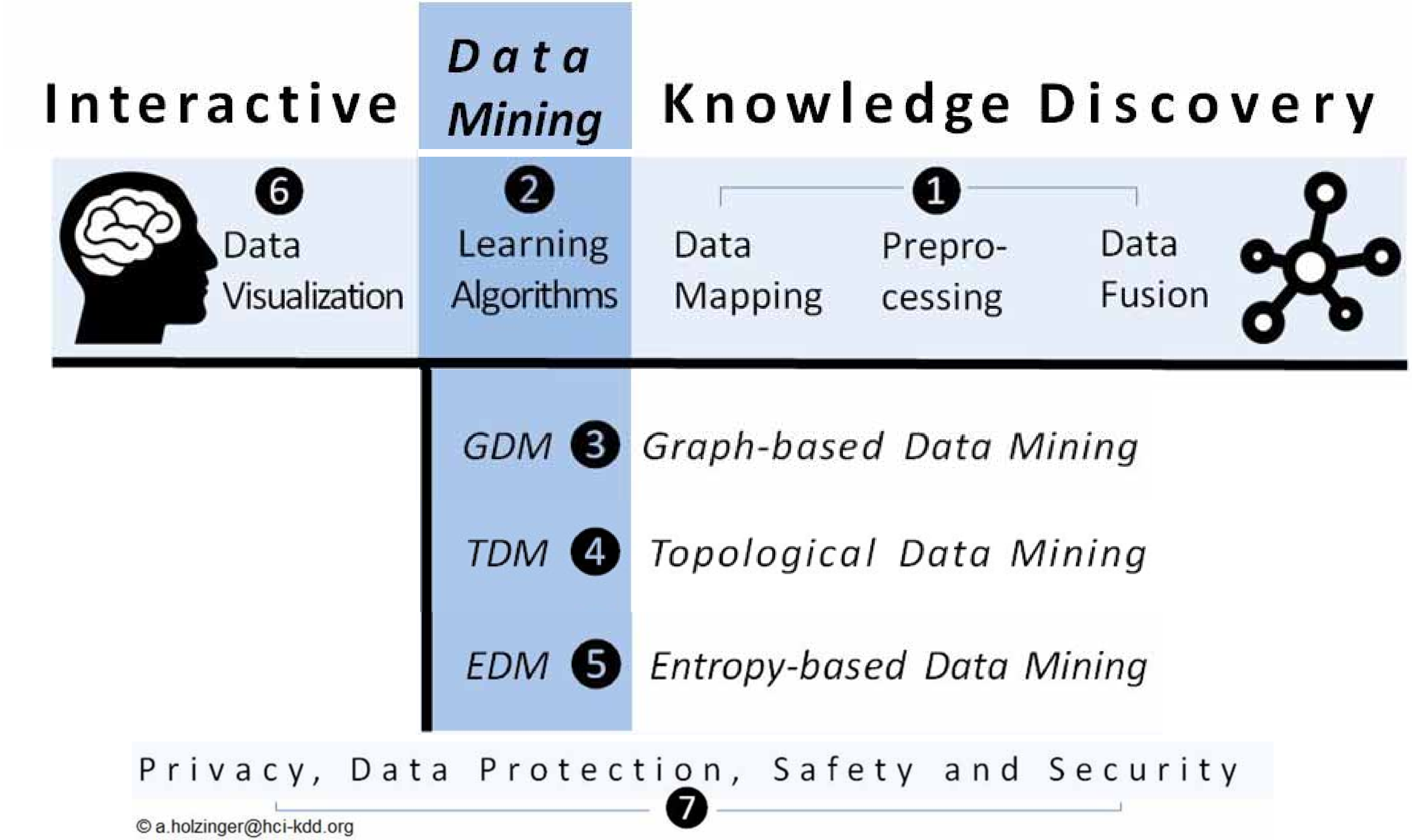
Transfer learning (TL)

is not easy: learning to perform a task by exploiting knowledge acquired when solving previous tasks:

a solution to this problem would have major impact to AI research generally and ML specifically.

Multi-Agent-Hybrid Systems (MAHS)

To include swarm-intelligence and crowdsourcing and making use of discrete models – avoiding to seek perfect solutions – better have a good solution < 5 min.



Holzinger, A. 2014. Trends in Interactive Knowledge Discovery for Personalized Medicine: Cognitive Science meets Machine Learning. IEEE Intelligent Informatics Bulletin, 15, (1), 6-14.

concerted effort

international

without boundaries …


HCI-KDD
You

HCI-KDD

Thank you!