# Thesis Title:

## Online fMRI Decoded Neurofeedback to Explore Metacognition and Awareness

**Name** Aurelio Cortese
**Course Name** International Priority Graduate Program, Information Science Faculty
**Professor** Prof. Kazushi Ikeda

**Abstract**

A central controversy in current studies of conscious perception concerns whether confidence judgements directly reflect the reliability of a perceptual - or cognitive - process, as suggested by Bayesian models, or alternatively, whether they are driven by other factors, such as subjective phenomenology. The Bayesian view enjoys popularity in both the computational and animal literatures, and stipulates that confidence and perceptual evidence share common neural substrates. The alternative view portrays confidence as dependent on a late-stage estimation process, itself dissociable from the actual reliability of signals. This second possibility would support the higher-order view of conscious perception as well as the global workspace theory. Yet, at least in humans, experimental tools have so far lacked the resolution to untangle these issues convincingly. Here, we overcome this challenge by employing a combination of machine learning, mathematical modeling, functional magnetic brain imaging (fMRI) and psychophysics. We first demonstrate that the neural representations of confidence, perceptual content and correctness of behavioral responses are only shared to a limited extent, particularly in frontoparietal areas. Then, through the recently-developed method of decoded neurofeedback (DecNef), we systematically manipulated multivoxel correlates of confidence in a frontoparietal network to up- and downregulate confidence in perceptual decisions in naive participants. Importantly, this was done without the participants's knowledge of the purpose of the neurofeedback procedure. By applying nonlinear mathematical modeling, we show that in sequential training sessions within the same participants, both increases and decreases of confidence are possible. Specifically, we demonstrated preserved confidence levels after an interval of one week as well as a weaker effect of DecNef in the second week. Our results provide clear neuroscientific evidence that confidence can be dissociated from perceptual performance. Further psychophysical analyses ruled out accounts based on simple shifts in criterion or reporting strategy for confidence. These findings challenge the normative views and support the late-stage model of confidence and metacognition. Furthermore, these results are interpreted in the context of reinforcement learning, and they provide important constraints on real-time multivariate fMRI applications to basic neuroscience as well as therapies for cognitive and mental disorders.

| 氏　名 | Aurelio Cortese |
|---|---|

（論文審査結果の要旨）

意識下の認知においては，確信度（confidence）が直接認知過程に影響をあたえるのか，それとも確信度は別の要素に駆動されたものなのかは明らかではなかった．本研究は，機械学習，数理モデル，fMRI 実験および心理物理学を組み合わせることで，この問題を解決したものである．

まず，前頭頭頂皮質における神経表現では，確信度，認知過程および行動実験の正確性は，共通している部分は限定的であることを示した．さらに decoded neurofeedback(DecNef) の手法を用いて確信度を外部から操作可能であることを示した．これは，確信度が行動実験の正確性とは別のものであることを意味している．

以上の実験の結果は確信度は別の要素に駆動されたものである可能性を支持しており，強化学習の枠組みでも説明可能である．これは fMRI を用いた DecNef の応用，特に神経科学研究および精神疾患治療において，考慮すべき重要な事項であり，今後の研究に大きな影響を与えるものである．

以上をまとめると，本論文は機械学習，数理モデル，fMRI 実験および心理物理学を組み合わせることで確信度の脳内表現を明らかにした研究であり，今後の神経科学の基礎および応用研究に大いに資すると考えられる．よって，博士（理学）の学位に値するものと認められる．