**Intuition & Reason: Re-assessing dual-process theories with representational sub-activation**

James Trafford[[1]](#endnote-1), Alexandros Tillas[[2]](#endnote-2) \*

**Abstract**

There is a prevalent distinction in the literature on reasoning, between Type-1 processes, (fast, automatic, associative, heuristic and intuitive); and Type-2 processes (rule-based, analytical and reflective). In this paper, we follow up recent empirical evidence [De Neys (2006b); Osman (2013)] in favour of a unitary cognitive system. More specifically, we suggest that intuitions (T1-processes) are sub-activated representations, which are in turn influenced by the weightings of the connections between different representations. Furthermore, we explain biases by appealing to the role of attention in thinking processes. The suggested view explains reasoning and bias whilst dealing with extant problems facing dual-process accounts.

[\**This paper was fully collaborative; the order of the authors' names is arbitrary*].

**1. Introduction**

**1.1 The current approach**

In the literature on reasoning, there is a prevalent distinction made between two processes, both of which underlie cognition, and often conflict with each other. Type-1 (T-1) processes are characterized as fast, automatic, associative, heuristic and intuitive; whereas Type-2 (T-2) processes are rule-based, analytical and reflective [Epstein (1994); Evans (1989); Evans & Over (1996); Sloman (1996); Stanovich (1999)].

The motivation for this distinction is, in large part, an attempt to understand the distinction between intuition and reasoning. For example, there are well-known discrepancies between the two as seen in matching biases [Wason & Evans (1975)], and biases pertaining to probabilistic reasoning [Kahneman & Tversky (1982)]. The latter are exemplified by the now infamous ‘Linda problem’ [Tversky and Kahneman (1982)]. In such cases, around 90% of participants routinely and systematically give a response suggesting that the probability of a conjunction (A&B) is greater than one of its conjuncts (A), (i.e. P(A&B) ≥ P(A)). One response to such evidence is to suggest that a ‘representative heuristic’ is involved in participant’s judgments, which is generated from T-1 processes. This, it is suggested, is indicative of a ‘belief bias’ [Stanovich (1999)] across human reasoning. The bias is also suggested by experiments where subjects presented with an invalid syllogism are asked to assess its validity [Sá, West & Stanovich (1999)]. In cases where the conclusion is believable, 68% responded that it is valid. Similarly, there is evidence suggesting that subjects tend to reject valid arguments where the conclusion is unbelievable [Evans, Barston, & Pollard (1983)]. Furthermore, DPTs provide plausible explanations of interpersonal variance in reasoning, [Stanovich (1999); Stanovich and West (1997)]; and cross-cultural variation in reasoning [Norenzayan et al. (2002)].

DPTs deal with these results by suggesting that these cases involve bringing prior beliefs to bear upon reasoning as a result of fast, automatic, associative T-1 processes.[[3]](#endnote-3) These in turn, are thought to require ‘override’, or inhibition, by engaging slow, reflective, analytic T-2 processes [Stanovich & West (2000); Chen & Chaiken (1999)]. Many theorists have gone on to suggest that the two processes are distinct enough to be characterized as two cognitive systems [Epstein (1994); Evans (1989); Evans & Over (1996); Sloman (1996); Stanovich (1999)], even though DPTs are not homogeneous.

More specifically, T-2 processing is generally slow and sequential, and involved in general intelligence processing such as hypothetical thinking, mental simulation, and decision-making [Evans (2007; 2010)]. According to [Evans & Stanovich (2013)], hypothetical reasoning requires a ‘cognitive decoupling’, i.e. keeping real-world representations separate from representations of imaginary situations, such as planning, counterfactual reasoning and so forth. Cognitive decoupling is the central feature of T-2 processing [Stanovich (2009; 2011)]. Finally, according to DPTs working memory becomes engaged *only* in T-2 processing.

Unlike T-2, T-1 processes are largely autonomous in the sense that they do not require ‘controlled attention’ and in turn have minimal demands on working memory resources. Execution of T-1 processes is initiated *only* in a bottom-up manner in the sense that they are triggered on presence of the appropriate stimulus in the subject’s vicinity. Operation of T-1 is mandatory and independent from higher-level cognition, [Stanovich (2004; 2009a; 2011)]. Despite the fact that all T-1 processes share an autonomous nature, they should not be construed as similar – let alone identical – with regards to their neurophysiology and etiology.

**1.2 Issues with the Two-Systems approach**

Whilst there has been much consensus regarding the distinction between T-1 and T-2 cognitive processes, the suggestion that there are two distinct cognitive systems has recently come under fire, and has been rejected by many of its original proponents [Evans (2006); Stanovich (2009)].[[4]](#endnote-4)

Nonetheless, it remains central to this programme that biases result from intuitions that require override through reason and analysis, and only participants that can suppress their intuitive response can correct their mistakes in relation to a normative theory. This view is bolstered by experimental data suggesting that subjects that are primed by making the normative structure apparent more successfully return the normatively ‘correct’ response (by e.g. the use of Euler circles, [Sloman & Over (2003)]; or using ‘foreign’ terms to construct syllogism’s, [Sá, West & Stanovich (1999)].

However, recent work suggests that such a conception of top-down inhibition by T-2 processes is overly-simplistic: (a) recent research suggests that T-2 systems are also liable to bias [Evans (2006); Thompson & Evans (2012)]; (b) T-2 processes appear to often be engaged to post-rationalise intuitive responses, rather than to exert cognitive control [Evans & Over (1996)]; (c) the intuitive phenomenology or ‘gut’ feeling as well as the belief bias appears to remain after logical reasoning is engaged [Evans, Allen, Newstead & Pollard (1994)].[[5]](#endnote-5)

Resultantly, we identify (at least) four significant issues in the extant literature:

1. Dual process approaches suggest a traditionally rationalistic approach to theory of mind in which biased intuitions are ‘controlled’ by serialized, normative processes. This suggests that we have a standard normative theory (usually classical logic / probability theory) – but this is rather simplistic given the numerous logical systems on the market e.g. Paraconsistent logics, relevant logics, non-monotonic logic and so on. [Stanovich (2004)] and [Oaksford et al. (1996)] recognize these issues, but they are largely undeveloped.[[6]](#endnote-6) Moreover, the measurement of ‘deviance’ from rationality in the construction of experiments is typically measured against a classical normative theory.
2. The relationship between processes of reasoning seems vastly more multifarious and mutually supportive than the dual systems approach allows. For example, it looks like systematic reasoning processes may also distort and bias heuristic reasoning processes. Moreover, there is evidence that T-2 processes may be engaged to post-rationalise T-1 processes in relation to introspective reports on behaviour. As it stands, we have no device for determining which expressed beliefs would represent an override by engaged reasoning, and which are mere rationalizations of T-1 processes [Evans & Over (1996)].
3. The intuitive pull of heuristic reasoning processes that appears to exist long after reflection is not really dealt with by DPTs (i.e. they don’t get the phenomenology right). Further work on the interrelated nature of reasoning processes is thus required – there is a problem of control here i.e. conscious reasoning does not seem to be able to exert the kind of control suggested by the DP approach (e.g. [Kahneman & Frederick (2002)] for this suggestion; [Evans et al. (1994)] for the suggestion that reasoning does not remove belief biases).
4. The dual-systems account may be overly internalistic, and thus not give the external scaffolding of systematic reasoning processes its due. In this respect, the role of learning; formalisation; technologies (e.g. writing, math etc.) are not adequately understood within the framework (and this is relevant to Issue 1).[[7]](#endnote-7) There are also reasoning processes that appear to begin under conscious, slow, control, but later become automated, such as learning logic & mathematics (e.g. [Monsell & Driver (2000)] in relation to attention and motor control). This suggests that the distinction between non-biased processes as consciously controlled, as opposed to biased processes as unconscious and automatic, may be too simplistic.

In sum, these issues suggest that it is worthwhile revisiting the explanation of biases in terms of dual-processes by investigating the structure of intuition, and so, more properly understanding its relation with reasoning.

**1.3 DPTs & Empirical Evidence**

In this section we focus on specific concerns regarding DTPs, primarily following Osman’s [(2004), p. 1006] argument that the evidence often deployed in favour of the distinction between T-1 & T-2 is actually suggestive of single-system accounts. More recently, Osman [(2013)] has argued that the evidence to which DPTs often appeal simply support quantitatively – and not qualitatively – different processes. Furthermore, she argues that hypothetical thinking in causal reasoning is not necessarily an instance of T-2 processing, and that the definitions of automaticity that DPTs suggest require further refinement, in the interest of avoiding ambiguity. Here, we discuss recent empirical evidence in the literature against DPTs, with an eye to fleshing out a philosophical account of a unitary system. In this sense, our suggestions build upon arguments raised by [Osman (2013)] and [De Neys (2006b)], in order to systematize them in a philosophical account of a unitary cognitive system.

*Accuracy & WM Loadings*

De Neys [(2006b)] focuses on some of the necessary defining characteristics of T-1 processing such as automaticity and immunity to WM loadings, as well as accuracy of performance in tasks that are allegedly executed in virtue of T-1 processing. More specifically, participants in this study were classified in three groups (low, medium and high) with regard to their Working Memory Capacity (WMC). In assessing WMC, participants were asked to accurately recall unrelated words while performing mathematical operations. (Given that both of those tasks load WM, the higher the WMC of a subject, the more words she would recall). Participants were then asked to evaluate a series of syllogisms. The believability of the conclusion of each syllogism was either incongruent with the logical conclusion (conflict syllogism) or congruent with it (no-conflict syllogism). While participants were evaluating the syllogisms, their WM was manipulated by using a dot memory test (high-, low-, no-loading). In particular, after evaluating a syllogism, subjects were asked to replicate a dot-array in a 3X3 grid that they saw prior the syllogism.

*Qualitative or Quantitative differences between T-1 & T-2 processes?*

Given the aforementioned working definitions for T-1 and T-2 processes, an intuitive prediction of DPTs would be that no-conflict syllogisms are processed in terms of T-1 processing. The underlying hypothesis here is that syllogisms with cohering premises and conclusion are easily integrated and evaluated, and subjects exhibit a tendency to accept the syllogism as valid (belief bias effect; [Wilkins (1928)], in [Osman (2013)]. Therefore, syllogisms of this kind should be easy-to-evaluate, and not load WM.

De Neys [(2006b)] shows that most often subjects tend to evaluate syllogisms by evaluating the believability of the conclusion. The results showed that accuracy levels, with respect to validity, were near ceiling, regardless of WMC and WM loadings. In particular, no-conflict syllogisms posed no WM loadings, and accuracy levels, also with respect to validity, were near ceiling [De Neys (2006b)]. According to Osman [(2013)], this suggests that T-1 processes underpin such evaluations.

On DPTs, we would also expect that responses in conflict tasks are generated in virtue of T-2 processing. This, presumably, is due to conflict syllogisms being hard(er) to evaluate, since when logic and belief disagree, subjects tend to base their responses on the believability of the conclusion. In evaluating conflict syllogisms, subjects integrate the premises and conclusion. This is a rather hard task since this integration is in contrast with the subjects’ existing world-knowledge [Klauer, Musch, & Naumer, (2000)]. In turn, integrating premises and conclusion has an effect of WM loading, [De Neys, (2006b)]. According to De Neys, regardless of the WMC of participants evaluating the syllogisms, accuracy levels was severely affected as the WM load increased. Given the defining characteristic of T-2 processing, it seems that responses were generated by T-2. Briefly, in line with the DPTs claims, simple tasks (those which do not load WM, and do not affect accuracy levels) are performed using T-1, whereas hard(er) tasks (those which load WM and affect accuracy levels) are executed by T-2 [Osman (2013)].

In contrast to DPTs, Osman argues that the above results do not indicate a qualitative difference between the two kinds of processes but merely suggest a quantitative distinction along a single dimension, i.e. the difficulty of the task. Similarly, it could be argued that participants are doing the same thing – call this belief management – in both cases of conflict and non-conflict syllogisms. Simply, in the case of valid syllogisms with believable conclusions (or invalid syllogisms with unbelievable conclusions), the ‘logical response’ and the ‘belief management’ response coincide. In turn, the issue of integrating premises that are in conflict with real-world knowledge is rather a different one, which also plays a role in the way reasoning tasks are processed.[[8]](#endnote-8)

*Speed of T-1 processes & immunity to WM loadings*

A further defining characteristic of T-1 processes according to DPTs is that they generate faster responses when compared to responses produced by T-2 processes. For, T-1 processes are allegedly autonomous, and in turn faster (than T-2), since they are immune to WM loadings, which naturally slow down reasoning processes [Posner (1978)], in [Osman (2013)]. (See also below for the relation between WM and attention).

De Neys [(2006a)] examined the immunity of T-1 processes to WM loadings by giving subjects the following selection task. A set of four cards was presented alongside a conditional statement such as an indicative or a deontic task, and subjects were asked to pick out cards that would test the rule.[[9]](#endnote-9) Performance accuracy levels in deontic tasks are supposedly much higher than in indicative tasks since prior knowledge about real world infractions of regulations facilitates responses. Furthermore, deontic tasks are allegedly executed in virtue of T-1 processing. Thus, Osman [(2013)] suggests that, according to DPTs, we can predict that WM loadings should have little or no effect on accuracy levels in deontic tasks since they are supposed to be generated by T-1 processes, which are immune to WM loadings. There should, however, be an effect of WM loadings on accuracy levels in indicative tasks.

In contrast to what DP theorists would intuitively predict, De Neys [(2006a)] shows that performance accuracy levels in both deontic and indicative tasks were significantly compromised when WM was loaded. As a consequence, Osman argues that either reasoning in deontic tasks is not generated by T-1 processes, or that T-1 processes are not autonomous. This, in turn, is in direct opposition to the defining characteristics that DPTs ascribe to T-1 processes. Based on this evidence, critics of DPTs argue that there are not two qualitatively distinct kinds of processes but rather one that generates accurate responses to simpler tasks and inaccurate responses to harder tasks.

*How automatic is automatic?*

The final line of criticism against DPTs that we consider here concerns the ‘automatic’ nature of T-1 processes, which, as explained, is closely related to the generation of rapid responses. De Neys [(2006a)] examined the response speed of subjects on deontic and indicative tasks, and recorded latencies for time spent reading the instructions and time spent making inferences.

The recorded results show that under WM load conditions, participants took about 26sec. to make an inference for deontic tasks vs. 20sec. for indicative tasks. Under no WM load conditions, subjects spent approximately 22sec. before making an inference in deontic tasks, and 20sec. for indicative tasks.

Comparison of times spent making inferences between both versions of tasks, and regardless of load manipulation, showed little or no difference, [De Neys (2006a); Roberts & Newton (2001)], in [Osman (2013)]. Clearly, this is in sharp contrast with the DPTs’ claim that T-1 processing solves deontic tasks, and so responses to such tasks are faster. Regardless of how response times in deontic tasks compare to those of indicative tasks, 26 sec. is seemingly too long for a process to qualify as automatic.

*What is automatic in T-1?*

Relatedly, De Neys [(2006a)] also recorded judgment times of responses to the conjunction fallacy (Linda problem). Typically in the DPTs literature, T-1 processes are thought to generate erroneous responses, given that the information presented in the description of the task automatically triggers prior knowledge. In turn, prior beliefs are deployed while making judgments about the likelihood of various statements (e.g. whether Linda is a bank teller and active in the feminist movement), that conflict with the correct response (according to classical probability theory).

De Neys reports that it took participants approximately 47sec. (excluding time spent on reading the instructions) to make the erroneous judgments, and approximately 57sec. to make the correct judgments. Despite that generation of correct responses is, on average, longer than erroneous responses, again 47sec. seems overly long for an automatic process. Recall that participants spent, on average, 19sec. to generate the accurate response in deontic selections tasks. Recall also that reasoning in deontic tasks (as well as the erroneous responses in tasks like the Linda problem) is allegedly generated by T-1 process. Thus, it seems that the speed of T-1 processes can either vary significantly from task to task (47sec. vs. 19sec.), which is certainly not a characteristic of an autonomic process, or that automaticity and speed of response is a task-relative measurement.

Having the above evidence and criticisms against DPTs in place, as well as Osman’s [(2004; 2013)] suggestions as a background setting, we move on to flesh out a philosophical account of how a unitary system operates. In what follows, we put forth a suggestion in the form of the Representational Sub-activation Thesis (RSA), according to which intuitions (or T-1 processes) are sub-activated representations, which are in turn influenced by the weightings of the connections holding between different representations. These structures of representations are associatively conditioned, they are largely network-like, and linked with other mental states such as beliefs and emotional states. Unlike DPTs, we argue that biases can be explained by appealing to the role of attention in thinking processes. As such, we suggest that the RSA framework both illuminates and provides explanation of reasoning and bias that may also deal with the extant problems facing DPTs. In this sense, we show that the suggested approach (RSA) accommodates the above evidence and explain the various ways in which it departs from DPTs.

**2. RSA as a Unitary System**

In presenting our philosophical account of how a unitary cognitive system operates, we suggest that intuitions (or T-1 processes) are sub-activated representations; the stronger the associations between a given set of representations, the stronger the appropriate intuitive pull. Furthermore, we argue that we can understand biases by appealing to the role of attention in thinking. More specifically, selectively attending to different aspects of a perceived stimulus, e.g. an invalid syllogism (see also §3.2 below), inevitably engages different cognitive strategies. Crucially, selective attention is driven by sub-activated stored representations in a top-down manner. The suggested view of a unitary cognitive system avoids the challenges that DPTs face.

**2.1 On the Nature of Intuitions: Background Commitments**

In fleshing out our view about the nature of intuitions, we appeal to a rather weak version of representationalism. We simply argue that we form perceptual representations through experiences with the world, and that these representations are stored in our minds. Also, we are committed to a view of concepts as structured entities built out of perceptual representations.[[10]](#endnote-10)

It is worth clarifying at this point that the suggested view builds upon an empirically vindicated version of Neo-Empiricism, (e.g. [Prinz (2002); Barsalou (1999)]), and departs from traditional Empiricist accounts of concept learning. We have developed a detailed Empiricist account of concept acquisition elsewhere [(Tillas (2010); (forthcoming 2015a)], where we explain how different kinds of concepts, including concepts with heterogeneous instances like danger; fuel and food, as well as abstract or lofty concepts like democracy and justice, could derive from perceptual experiences. These kinds of concepts have been notoriously hard to account for in terms of traditional Empiricists accounts.[[11]](#endnote-11) The key point here is that all concepts are linked back to experience temporarily, but they do so *indirectly*.[[12]](#endnote-12) In this sense, the causal relation between mind and world is a fairly sophisticated one. For our present purposes it suffices to stress that in the light of perceptual experiences with instances of a given kind, the connection weightings of neuronal groups that ground or underlie these experiences are adjusted. At this point we appeal to the ubiquitously accepted principles of Hebbian learning (see below). Hebb’s [(1949)] rule of learning is vindicated by contemporary evidence for Long Term Potentiation (LTP).[[13]](#endnote-13) Having this empirically vindicated empiricist view as a starting point, we suggest a view that avoids the pitfalls of traditional views while maintaining the perceptual priority thesis according to which everything that is in the intellect is first in the senses.

Admittedly, many views hold that concepts are (or are tokened by) mental symbols; for the most part, our suggestions can be reformulated in those terms, so long as such views can deal with the idea that mental symbols are connected with adjustable weightings.

Furthermore, we suggest a way in which concepts may be built from perceptual representations, and focus on the relations between these representations.[[14]](#endnote-14) Finally we argue that the relations between different representations and concepts are weighted in certain ways, and that these weights constrain thinking. As such, intuitions can play a role similar to conscious thoughts in terms of how thinking is constrained.

**2.2 Perceiving the World & Building Networks**

Recall our commitment that on perceiving a given object a representation is formed and stored in long-term memory. The precise locus for storing a given representation is influenced, and often determined by top-down effects from stored representations in the subject’s mind.[[15]](#endnote-15) We have elaborated on this issue elsewhere [Tillas, (2014); (forthcoming 2015b)]. Briefly, on encounter with a subsequent instance that does get attended and recognized as a subsequent instance of a given kind, a representation is formed. At this point, a scanning process gets initiated and a match is sought for in the subject’s memory. The same process was also initiated during encounter with the first instance but the scan did not yield any matching stored representations.

The scanning process begins with highly demanding similarity level, which drops if a match is not found after a scan. If the similarity level drops below a certain point (after a certain number of scans), a new ‘mental file’, for the new category, is formed, (e.g. [Perry, (2001)]). The location of where a given representation will be stored is not only influenced by similarities with existing representations but also by contextual features as well as other known information about the current encounter.

Once a match between the currently formed and some/a stored representation(s) is found, then the existing representation(s) are/is activated.[[16]](#endnote-16) The activation of this representation leads to two things. First, stored representations lead perception in a top-down manner. For instance, if on perception of the first instance of a tree, a subject has formed a representation of the tree’s trunk, and a match with a stored representation is found, then selective attention will be driven to the rest of the tree parts (e.g. branches, leaves) that have been attended during the encounter with the initial instance. Second, activation of matching stored representations leads to storage of the currently formed representation at the same locus or ‘closer’ to stored matching representations.

A possibly useful way to construe the claim that representations are stored ‘closer’ together might be in terms of representations with stronger positive memory effects in cognitive tasks. This hypothesis builds on evidence showing that unlike bottom-up attention, i.e. attention captured by salient features, top-down attention enhances formation of representations of attended features/aspects/information e.g. [Corbetta et al. (1990); Noudoost et al. (2010)], as well as on evidence showing that information attended through top-down attention will later on be relevant for memory formation [Uncapher et al. (2011)] or later remembering [Craik et al. (1996)].

It is worth clarifying that the cognitive system does not rely merely upon top-down effects of stored representations on perceptual processes. Had that been the case, the perceptual process would have been heavily influenced by accidents of the first encounter with a certain instance and miss out on certain statistical regularities. To this end, top-down influences have to be construed more liberally, in the sense that some non-overlapping information is allowed as a by-product of not being too skewed by the bias of the first encounter.

On encounter with subsequent instances of a given kind, a similar process will be initiated, and the appropriate mental-file will get informationally enriched.

Digressing slightly for a moment, given that the suggested views builds heavily upon the principles of associationism as well as top-down and bottom-up influences it naturally converges with what is known in the literature as ‘predictive coding’ (e.g. [Clark (2013)]). On that view, natural signals are highly redundant due to spatial and temporal uniformity. For instance, intensities of pixels tend to be correlated over time given that objects persist in time. Given these redundancies in natural signals, the suggestion is that representing a raw image directly in terms of the activity of a set of sensory receptors would be very inefficient. Thus, it is argued that early sensory processing aims at reducing such redundancies and recoding sensory inputs in more efficient ways. Predictive coding postulates that neural networks learn the statistical regularities inherent in external stimuli, and reduce redundancy by removing components of the input that are predictable. In turn, they will transmit residual errors in prediction. In the case of visual perception, for example, cells in early visual circuits convey not the raw image intensity, but the difference between the predicted value and actual intensity.[[17]](#endnote-17) From this brief exposition, it should be clear that even though predictive coding operates at a different level to the present view, it should also be clear that the present view is in line with predictive coding. For in the present view, connection-weighting calibration (which captures the dynamic character of representational networks as well as key aspects of thought production) is construed as governed by top-down and bottom-up influences as well as by the principles of associationism and statistical learning.

**2.3. Co-occurrence and Network Dynamics**

In this section we focus on the relations between different representations, concepts and thinking. In doing so, we appeal to Hebb’s [(1949)] famous hypothesis about the ways in which neurons are changed by experience in such a way that this change would have altering effects in behavior. The widely accepted Hebbian rule ‘cells that fire together, wire together’ is a type of correlational learning rule in which the temporal coincidence of pre- and post-synaptic activity results in a strengthening of that synapse. Thus, whenever two neurons get excited simultaneously, the connections between them get strengthened. According to Hebb, repeated participation of neuron A in firing neuron B entails an increase of the strength of the action of A onto B. That is, on the basis of repeated simultaneous activation, the efficiency of one cell to activate the other is increased.

Recall the process of perceiving the first instance of a tree. The subject selectively attends to different parts of the perceived tree and representations of those parts are formed and stored in the mind. This process is underlain by neuronal activations. Given that different parts of the perceived tree appear different, are in different positions, etc. and also given that certain neurons are dedicated to perception of edges, colours etc., the claim put forth here is that each part of the perceived object is represented on the basis of different neurons firing in a particular way. Note that we do not imply a one-to-one mapping between neuronal activations and features in the world.

According to aforementioned Hebbian rule, in order for different cells to wire together, they have to get activated *simultaneously.* In our view, perception does occur in virtue of ‘simultaneous’ firings of different neurons dedicated to perception of specific aspects and features of perceived object and events. More specifically, visual selective attention is by and large dependent on saccadic eye movements. Given that each saccade lasts a few milliseconds, we argue that perception of different parts of an object, of a tree in this case, occurs (almost) simultaneously. In turn, properties of different instances that get systematically co-activated will get strongly connected.

At a superordinate level, concepts, like single representations, get also connected with each other. The more frequent the co-occurrence of a pair (or set) of concepts the stronger the connections between them. In line with the Hebbian rule, once a concept of this pair becomes activated, the other will also become activated. Clearly, selective attention might override the aforementioned Hebbian connections between two concepts, but without such influences, the Hebbian principle applies also to concepts. Crucially, the rest of the (stronger) connections stemming from an activated concept also contribute in activating different other concepts. [[18]](#endnote-18)

On the hypothesis that concepts are the building blocks of thoughts, the connections between concepts will influence or determine (though not necessarily in a systematic way), which thoughts will next be formed. We turn to examine this process in more detail next.

**2.4. Thinking about Thinking**

As is standard, concepts are taken to be the building blocks of thinking. In turn, a crucial aspect of concepts is that they have rich cognitive role properties. More specifically, we focus here on the suggestion that concepts are endogenously controllable or that can be activated in a top-down manner (e.g. [Prinz (2002); Barsalou (1999)]). [[19]](#endnote-19) In this sense, concepts and in turn thoughts can be formed not only on the basis of processes of perceiving but also of processes of thinking.

In the suggested view, endogenously controlled thinking is a form of associative thinking; that is, current thinking caused by earlier thinking. Here we commit to a view of internal thinking which is imagistic, to the extent that conceptual thoughts are built out of concepts, which are in turn built out of perceptual representations. In turn, we take concepts to be associationistic in their causal patterns. That is, every concept is associated with other concepts (semantic content). Once a given concept is activated, concepts associated to it will also be activated. For example, consider someone uttering the word ‘Trip’ and another agent mistakenly hearing the word ‘Grip’, who may resultantly start to think about friction and laws of physics, rather than travelling. This is a case where an agent is forming a thought in the absence of an appropriate stimulus, seemingly in a spontaneous but actually in an associative manner.

Evidence in support of the suggested associationistic view of thinking can be found in the work of Elman et al. [(1996)], even though we are not committed to a connectionist cognitive architecture. According to Elman, artificial neural networks can be highly constrained by the network’s current weight assignment. That is to say, that the pattern of activation set by a connectionist network is determined by the weights, or strength of connections between the units. These weights model the effects of the synapses between different neurons in the human brain. So, different levels of activation of synapses that connect one neuron to another place a significant constraint on what new ideas the mind can explore next. In this sense, sub-activation of certain neuronal ensembles constrain or drive activation in a specific way, i.e. towards specific thoughts associated with the sub-activated links of the linguistic network in question.

Given the current understanding of intuitions (or T-1 processes) as largely unconscious cognitive processes, which are projected in to conscious cognitive processes, we argue that intuitions are precisely these sub-activated connections between different concepts. These, in turn, are influenced by the relative weights between their associations. In line with Hebbian associationism, the more frequent the co-occurrence between different concepts, the stronger the connections between them becomes. In turn, the stronger the connection between sets of concepts, the more frequently certain sequences of thoughts tokened, and hence, the more intuitive it will seem to us that a certain given thought will entail a certain other thought. Once again, intuitions are these sub-activated representations, which influence thinking.

**3. Sub-activated Representation & Intuitions**

**3.1. Connections with the Dual-Process account**

In the previous section, we suggested an account about how thinking occurs (call it Representational Sub-activation Thesis, or RSA) and put forth a view concerning the nature of intuitions. In this section, we focus on the way that the suggested view relates to DPTs, and how it deals with some of the extant issues with those approaches, identified in (§1.2).

RSA relates to DPTs roughly as follows. We take intuitions, or T-1 processes, to be sub-activated representations, which are influenced by the relative weightings of the connections that hold between different representations. These structures of representations are largely network-like, and representations within these networks are linked with other mental states (e.g. emotional states), and are associatively conditioned.

Fleshing out this idea, the stronger the weight of the connection between a pair of representations, the greater the probability that activation of one representation will activate the other (Hebbian rule). As it happens, a given representation in a network is further associated with a number of other representations. Once the representation in question is activated, the rest of the associated representations will become sub-activated. The reason why not all associated representations become activated, but rather sub-activated, is two-fold. First, different weights obtain across different sets of associated representations. In addition, the strengths of the connections determine the levels of activation. Importantly, which representations will be fully activated is also influenced by selective attention. Recall from the previous section that stored sub-activated representations drive selective attention in a top-down manner. In this way, once a stored representation is activated, the associated representations are sub-activated, and, so drive selective attention to certain parts of the stimuli or to other stored representations (in the case of off-line thinking). In turn, once selective attention focuses on a further representation within the set in question, a similar process occurs where there are top-down influences on selective attention and the sub-activation of further associated representations. In this way, the mind explores, or forms, sequences of ideas. Admittedly, the above-described process is merely an idealised scenario of off-line thinking, since in many cases of online (and to a lesser extent off-line thinking processes) further factors, such as contextual features, influence selective attention.

It is worth noting that the same representational network-like structures are also available to reasoning (T-2) processes. In this sense, reasoning processes may also become activated in a top-down manner via directed endogenous control. So, thinking is constrained by the weights of different associated sets of representations. The weights of different associations are, in turn, determined by frequencies of the co-activation of representations, whether emotional states are involved in those processes, and so forth.

In this sense, representational structures enjoy great levels of plasticity, in the sense that in light of new stimuli or reflection, new associations can be formed and, in turn, direct the system in various ways. The more frequently attention is drawn to a given new representation the faster an association will be formed.[[20]](#endnote-20) In contrast, the more rarely an existing association is activated, the weaker the connection between them will grow, and the more idle it will become in thinking processes. It is in these terms that we might consider the process of debiasing, and more fully explain fundamental biases such as the belief bias.

**3.2 Explaining “bias”**

Consider a classic case intended to display “belief bias” in thinker’s reasoning. Here, belief bias has to do with subject’s tendency to endorse the validity of invalid arguments that accord with extant beliefs, and reject valid arguments that do not. The bias is highlighted in experiments [Sá & Stanovich (1999)], where subjects presented with an invalid syllogism, like the following, are asked to assess its validity.

1. All living things need water.
2. Roses need water.
3. Thus, roses are living things.

According to the obtained results, 32% of subjects replied that this was not a valid argument, while 68% returned valid. The mood of the syllogism is AAA-2. A syllogism of the same mood was then presented to the subjects involving ‘foreign’ terms (involving an imaginary species, Wampets, and an imaginary class, Hudon), and subsequently, they were asked to evaluate the following:

1. All animals of the Hudon class are ferocious.
2. Wampets are ferocious.
3. Thus, wampets are animals of the Hudon class.

Here, 78% of the subjects gave the “logically correct” response here, that the syllogism is invalid. According to proponents of DPTs, subjects seem to bypass logical principles that would otherwise allow them to identify the syllogism as invalid. In this sense, DP theorists appeal to a hypothesized dichotomy between intuitions and reasoning, where reasoning is in broad accord with classical logical principles.

However, as pointed out in §1.2, we suggest that this model is overly rationalistic, and poorly reflects the role of normative reasoning strategies. In general, the suggestion that “biased” responses to the first syllogism requires a normative theoretical system to be *already in place* in the thinker’s cognitive architecture, which is thought to follow the laws of classical logic or probability theory. Given the substantive disputes regarding logical systems that have emerged over the last thirty years or so (concerning, paraconsistent logics, relevant logics, non-monotonic logics, and so on) this appears overly parochial. Resultantly, we prefer to employ the phrase “doxastic conservativeness” [Dutilh Novaes (2012)] to characterise the tendency of reasoners to bring to bear prior beliefs on the assessment of argument.

RSA provides a promising explanation that models doxastic conservativeness in this experiment as follows. Recall, from §2, three key elements of the RSA model that play a crucial role in explaining the obtained result from the invalid syllogism experiment.

* Thinking is influenced by connection weights between associated representations.
* RSA appeals to the role of selective attention focusing on aspects of the perceptual stimuli.
* Selective attention is driven in a top-down manner by strongly connected stored representations.

Consider the first syllogism. According to DPTs, subjects return the logically false reply because of an intuitive pull, which overrides rational processes. In contrast, we suggest that on confrontation with the syllogism in question subjects attend to certain aspects of premises 1-3. If not seen as a syllogism in which 3 is supposed to follow from 1 and 2, i.e. merely by ignoring the predicate ‘Thus’ in (3), then it is rather a list of sentences (as a matter of fact, it is a list of truisms). Given prior beliefs, involving associations between representations of roses and living things, there is a match between the input-sentence “Thus, all rose are living things”, and representations of roses being associated with living things. Given this matching, the subject will feel the intuitive pull to confirm the validity of the syllogism. This, as explained, is due to the automatic representational sub-activation of the association between token representations of roses and living things.[[21]](#endnote-21) Subjects cannot simply disregard the representational content of the words “rose” and the phrase “living things” because of this underlying process.[[22]](#endnote-22) This explains why we display the tendency to make judgments based on extant networks of representations involving doxastic states, emotions, and other mental states.

Had it been the case that the subjects have actually treated the syllogism as a syllogism rather a list of truisms, in the sense explained above, then the same matching would have occurred with representations or stored knowledge about certain logical rules, universal quantifiers, etc. (if the subject was trained in logic). In turn, the intuitive pull would have been against the truism that all roses are living things, and the focus would have been on “Thus” and in turn on the fact that (3) does not follow from (1) & (2). [[23]](#endnote-23) In line with what has been said in the second part of the paper, once there is a matching with stored representations, attention will be driven in a top-down manner to search for existential qualifiers in the premises/steps of the syllogism. Given the way the syllogism in question is couched, the intuitive pull to resist the conclusion will be even stronger. This is precisely what we find in the second example, where the syllogism is couched in “foreign terms” that do not activate automatic RSA processes in the same manner, because of lack of representational associations.

Note here that if the subject is not acquainted with norms of classical logic, it might well be the case that they will still accept the invalid conclusion as true. For in this case, there will be no stored representations to allow the subject to resist the conclusion – in essence, this latter case will greatly resemble the original one. Note also that most often it takes some time to form associations between representations (but see footnote 12). It is for this reason that sometimes the intuitive pull for accepting that the invalid conclusion as true, for instance, lingers on even after it has been explained to the subjects that this syllogism contradicts the norms of classical logic. Hence, there is no need to invoke a dualistic cognitive architecture here that assumes the canons of classical reasoning.

In the second case, there will not be an intuitive pull towards accepting the syllogism as true since unlike the previous case of the syllogism couched in terms of natural language that the subject understands, there will not be a priming of certain representations. Moreover, if the subject is familiar with certain logical rules, there will be a matching between the inputting representation and stored ones, since the formalization will sub-activate representations associated with logical rules, and so on. In turn, the subject will be primed in a top-down manner to selectively attend to the structure of the syllogism.

The above suggestions also explain why so-called ‘T-2’ processes are themselves susceptible to “biases”, precisely because they are not distinct cognitive processes. As it happens, there is only one kind of reasoning process, which just happens to occur by deployment of different sets of representations in each of the two cases (formalised and non-formalised syllogism). In this sense, awareness of one’s own reasoning processes can increase bias through post-rationalisation of the RSA process.[[24]](#endnote-24)

As already noted, most often it takes time in order to form associations between representations, and, in order to do that, the subject needs to consciously attend to specific aspects of the stimuli. For instance, learning the validity of syllogisms is a slow and progressive process. However, once a certain level of strengths between different representations is reached, the activation of one of those associations will trigger sub-activation (or full blown activation) of other associated representations. In this sense, the representations that required selective attention in order to get formed, stored and triggered can in time be activated automatically merely in virtue of one representation of the set in question being triggered. In turn, the more automatic this activation pattern becomes, the more intuitive their phenomenology – so to speak – becomes. Thus, the RSA account is in a position to account for the learning processes involved in, for example, critical thinking, logic, and mathematics. It is possible to see how a learning process will allow processes that begin slowly and consciously (such as learning the validity of syllogisms), to become automatic, since that process will effectively prime RSA to form new associations.

**4. RSA & empirical evidence**

In this section, we evaluate the RSA ‘model’ against the empirical evidence that we cited earlier against DPTs, and examine whether the suggestion that we put forth here can accommodate it.

Recall the aforementioned evidence from work on no-conflict tasks. De Neys [(2006b)] reports that performance was near ceiling and that there was no loading on Working Memory. DPTs predict these results by arguing that participants relied upon real world knowledge – in evaluating the syllogisms in question. In turn, relying upon world knowledge is significantly easier than implementing rules of logic in assessing a syllogism’s validity (since it uses less cognitive resources).

According to RSA, the knowledge upon which participants rely is captured by stronger associations. In this sense, there will be some loadings on WM but this will be less than when dealing with conflict syllogisms. Nonetheless, there will be some loading on WM regardless of the difficulty of the task (see below). The more practised subjects are in performing similar tasks, and in turn the closer to automaticity such evaluation processes get, the less taxing processing of similar tasks will become on WM.

Recall also the rather simple, empirically verified [De Neys (2006b)], DPTs prediction that T-2 processes underpin conflict tasks evaluation. For, conflict syllogisms are difficult to evaluate and, thus, there should be an effect of WM load. Even though it is plausible to assume that conflict syllogisms are harder to evaluate, we disagree with how DPTs explain this evidence. Recall that DPTs argue that when logic and belief disagree, subjects appeal to the believability of the conclusion. In turn, they will find it harder to integrate the premises and conclusion of a conflict syllogism since this integration challenges their prior world knowledge (conflict syllogisms).[[25]](#endnote-25)

In contrast, we argue that the difficulty in assessing conflict syllogisms stems from the principles of Hebbian learning. Briefly, the more frequently certain beliefs co-occur, the stronger the associations between representations that ground them. Given that conflicting ideas do not co-occur very often, the representations that ground them are only weakly connected – if at all. Thus, in cases of conflict syllogisms participants do not have the ‘disposition’ to infer the conclusion from the premises. Without this disposition in place to drive selective attention to the aspects of the syllogism that refer to principles of logic, e.g. structure of the syllogism, the word ‘Thus’, etc., they attend to words of the syllogisms that couch these conflicting beliefs. Thus, in essence, they bring to bear prior beliefs while assessing the syllogism in question (doxastic conservativeness). Given weak associations between these beliefs, participants end up evaluating in detail the associated semantics, which clearly is a cognitively taxing process.

* *WMC & accuracy levels*

Recall also that regardless of the WMC of participants evaluating syllogisms, the accuracy of their performance was severely affected as the load increased. At first glance, this runs against our claims that reasoning, as well as intuitions, is influenced by connection weights. For, this claim might be taken to imply that activation of strongly associated representations will not have to go through WM, and thus accuracy levels will be unaffected by WM loadings. The principles of Hebbian learning seem to imply that activation of neuronal group X will ‘automatically’ trigger neuronal group Y, and so forth.

However, RSA does not render reasoning immune to WM loadings. In fact, there are strong interdependencies between attention and WM.[[26]](#endnote-26) Let us elaborate. Recall that dispositions are analysable in terms of stronger associations between representations. One of the effects that these associations have on perception is that they can drive selective attention to certain aspects of the perceived stimulus, e.g., the structure of a syllogism and treat it as such instead of as a list of truisms. Crucially, in order for attended information to contribute to higher cognitive processes such as reasoning, this information has to be accessible by WM – (presumably this precondition might become less pronounced as a given process becomes automatic). Thus, despite the top-down urge to attend to certain aspects of the stimulus, if WM is loaded, the attended information will not be accessible/usable during evaluating the syllogism in question.

As we saw above, according to DPTs, subjects with high WMC should show an effect of WM loadings, given that they are using analytic processes, which typically slow and drain WMC. This prediction has not been empirically verified. Instead subjects with low WMC perform worse than those with high WMC, [De Neys (2006b)]. In line with the latter piece of evidence, we argue that subjects with high WMC can exploit attended-information to a greater extent when compared to low WMC subjects, simply because processing of attended information requires WM. Thus, it is not appropriate to treat WMC in a ‘use-more-if-you-can-afford-it’ manner. Rather, it is more plausible to assume that only the WM resources necessary for a specific task will be used/deployed, while the remaining WM capacity is standing by. Thus, we expect loading of WM regardless of WMC, even though subjects with higher WMC are (ceteris paribus) expected to be more accurate either in the selection task or the dot memory test or at both.

According to DPTs, accuracy levels in deontic tasks should not be affected by WM loadings. In contrast, WM loadings do have effects on accuracy in indicative tasks – this prediction has not been verified. Specifically, the reported results show that WM loadings have a negative effect on accuracy levels (poorer performance when compared to no WM loads) on both deontic and indicative versions of the selection task ([De Neys (2006b); Oaksford, Morris, Grainger, and Williams (1996)]). It thus seems plausible that reasoning in both deontic and indicative tasks requires WM. In turn, this contrasts with DPTs and is suggestive of a unitary cognitive system.[[27]](#endnote-27)

* *Why accuracy levels were higher in deontic tasks?*

According to RSA, accuracy levels in deontic tasks are higher – in comparison to indicative ones – given that deontic tasks concern well-known regulations. Intuitively, well-known regulations are underpinned by strongly interconnected representations. In turn, there will be top-down effects on perception of the deontic task, which will be processed in light of existing dispositions/biases. RSA predicts these results, given the role that it ascribes to prior experiences in reasoning.

An alternative interpretation of the performance results in deontic tasks comes from Stenning and van Lambalgen [(2004)]. Their prediction is that descriptive/indicative tasks will be highly problematical and deontic tasks rather straightforward. As shown above, this prediction is confirmed by the results. In more detail, their starting point is that the main interpretative problem facing subjects in reasoning tasks is assigning logical form to the task at hand or providing settings for all the involved parameters. The problem, they argue, is that psychology of reasoning has traditionally operated on an oversimplified notion of logical form. Namely, assigning logical form has been traditionally construed as translating a natural language sentence into a formal language with given semantics. Stenning and van Lambalgen expand this notion of logical form and argue that indicative/descriptive and deontic tasks are actually different in terms of logical form. With regards to the strategy subjects follow in reasoning tasks, they argue that most likely subjects do not really know what they are doing but they certainly worry about how to set the parameters of the task. Furthermore, they argue that different subjects adopt different strategies. In light of this focus on the differences in logical form between indicative/descriptive and deontic tasks, the interpretation of DPTs – who focus mainly on WM loadings – is rendered rather oversimplified. The view suggested here is ‘compatible’ with this alternative view to the extent that it does not focus simply on WM loadings as the main influence of performance at reasoning tasks. As explained above, our focus is (also) on the conceptual networks deployed in reasoning tasks. These networks might also be construed as carrying information concerning logical form. (See [Tillas (2010)] for a detailed discussion).

* *In contrast to RSA, results show that participants took longer to respond to deontic tasks (in comparison to indicative tasks). Naturally, given stronger associations, one should expect processing of deontic tasks to be faster than that of indicative ones.*

Stronger connections between representations are ultimately stronger connections between neuronal ensembles. In turn, stronger interneural connections allow more current to go through them and thus exhibit facilitatory effects in processing of a given signal. However, the more associated information deployed during a given task, the slower their processing would be, regardless of the speed of processing of specific bits of information. This is because processing of more information naturally requires more resources. Given that deontic tasks concern well-known regulations, it is likely that a significant amount of information will be associated with these regulations. At least, this associated information will be much greater in comparison to representations of arbitrary rules that the subjects encounter possibly for the first time. Admittedly, the more automatic processing of a given deontic task becomes, the less time a given subject will need to process the deontic task in question. However, this does not imply that processing of deontic tasks should in principle be faster than that of indicative ones.

* *Regardless of time needed to respond, accuracy levels in deontic tasks were also influenced by WM loadings.*

According to RSA, regardless of where representations deployed in reasoning come from (long-term memory in the case of deontic tasks & short-term in the case of indicative tasks), they need be placed in WM in order to feature in reasoning processes. Naturally, if WM is loaded, then this will have a negative effect on the way that the relevant information becomes deployed in reasoning. For instance, it might be that only a part of this information is actually processed in reasoning, while the rest is processed to a lesser extent or is simply ignored. Thus, loaded WM will naturally yield less accurate results.

* *Regardless of WM loadings, responses in deontic tasks were reported to be slower than indicative.*

In line with the above remarks, the reason why processing of deontic tasks takes longer is because information strongly associated with the information in the deontic task, will inevitably get activated. Given that deontic tasks concern well-known regulations it is more likely that a significant amount of representations is associated to the regulations the tasks concern. In contrast, there is no informational ‘luggage’ in indicative tasks. This is why indicative tasks may yield faster responses.

* *Instead of having two kinds of cognitive processes, it might be the case that the same reasoning process generates accurate responses to easier tasks and inaccurate responses to harder tasks, [Osman (2013)].*

This is also something that RSA predicts. Simpler tasks most often concern non-conflicting information, while harder tasks concern conflicting information. As explained, conflicting information are less likely to be associated with stronger connections, if connected at all. More specifically, harder tasks are tasks in which participant have to bring not-strongly-connected-information to WM (via selective attention) and process it. This inevitably leads to slower responses and lower accuracy levels.

**References**

1. AWH, E., JONIDES, J., SMITH, E. E., BUXTON, R. B., FRANK, L. R., LOVE, T., et al. (1999), ‘Rehearsal in Spatial Working Memory: Evidence from Neuroimaging’, *Psychological Science*, 10, 5, pp. 433-7.
2. BARSALOU, L. W. (1999), ‘Perceptual Symbol Systems’, Behavioral and Brain Sciences, 22, pp. 577-609.
3. BLISS, T. V. P., & GARDNER-MEDWIN, A. R. (1973), ‘Long-Lasting Potentiation of Synaptic Transmission in the Dentate Area of the Anesthetized Rabbit Following Stimulation of the Perforant Path’, *Journal of Physiology (London)*, 232, pp. 357-74.
4. BLISS, T. V. P., & LOMO, T. (1973), ‘Long-Lasting Potentiation of Synaptic Transmission in the Dentate Area of the Anesthetized Rabbit Following Stimulation of the Perforant Path’, *Journal of Physiology (London)*, 232, pp. 331-56.
5. BROWN, T. H., & MCAFEE, D. A. (1982), ‘Long-Term Synaptic Potentiation in the Superior Ganglion’, *Science*, 215, pp. 1411-3.
6. CHAO, L. L., HAXBY, J. V., & MARTIN, A. (1999), ‘Attribute-Based Neural Substrates in Temporal Cortex for Perceiving and Knowing About Objects’, *Nature Neuroscience*, 2, pp. 913-9.
7. CHEN, S. & CHAIKEN, S. (1999), ‘The Heuristic-Systematic Model in its Broader Context’, in Chaiken, S. and Trope Y. (eds.), *Dual-Process Theories in Social Psychology*, New York, Guilford Press, pp. 73-96.
8. CORBETTA M, MIEZIN FM, DOBMEYER S, SHULMAN GL, & PETERSEN S.E. (1990), ‘Attentional Modulation of Neural Processing of Shape, Color, and Velocity in Humans’, *Science,* 248, pp. 1556-9.
9. COWAN, N. (2006), *Working Memory Capacity*, New York, Psychology Press.
10. CRAIK, F. I. M., GOVONI R., NAVEH-BENJAMIN M. & ANDERSON N.D. (1996), ‘The Effects of Divided Attention on Encoding and Retrieval Processes in Human Memory’, *Journal of Experimental Psychology General*. 125, pp. 159-80.
11. D'ARCAIS, G.B.F. AND SCHREUDER, R. (1987), ‘Semantic Activation During Object Naming’, *Psychological Research*, Vol. 49, Issue 2-3, pp. 153-9.
12. DE NEYS, W. (2006a), ‘Automatic-Heuristic and Executive-Analytic Processing in Reasoning: Chronometric and Dual Task Considerations’, *Quarterly Journal of Experimental Psychology*, 59, pp. 1070-100.
13. DE NEYS, W. (2006b), ‘Dual Processing in Reasoning: Two Systems but One Reasoner’, *Psychological Science*, 17, pp. 428-33.
14. DEMARAIS, A. M., & COHEN, B. H. (1998), ‘Evidence for Image-Scanning Eye Movements During Transitive Inference’, *Biological Psychology, 49,* 3, pp. 229-47.
15. DUTILH NOVAES, C. (2012), ‘*Formal Languages in Logic: A Philosophical and Cognitive Analysis’,* Cambridge, Cambridge University Press.
16. ELMAN, J. L., & MCCLELLAND, J. L. (1986), ‘An Architecture for Parallel Processing in Speech Recognition: The TRACE model’. In Schroeder M. R. (ed.), *Speech Recognition*. Basel, S. Krager AG.
17. EPSTEIN, S. (1994), ‘Integration of the Cognitive and Psychodynamic Unconscious’, *American Psychologist*, 49, pp. 709-24.
18. EVANS J. (2008), ‘Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition’, *Annual Review of Psychology*,59, pp. 255-78.
19. EVANS, J. B. (1989), ‘*Bias in Human Reasoning: Causes and Consequences*’, Brighton, Erlbaum.
20. EVANS, J. B. (2006), ‘The Heuristic-Analytic Theory of Reasoning: Extension and Evaluation’, *Psychonomic Bulletin & Review*, 13.3, pp. 378-85.
21. EVANS, J. B. (2007), ‘On the Resolution of Conflict in Dual Process Theories of Reasoning’, *Thinking & Reasoning,* 13, pp. 321-39.
22. EVANS, J. B. & OVER, D. E. (1996), ‘*Rationality & Reasoning’,* Hove, Psychology Press.
23. EVANS, J. B., ALLEN, J. L., NEWSTEAD, S. E. & POLLARD, P. (1994), ‘Debiasing by Instruction: The Case of Belief Bias’, *European Journal of Cognitive Psychology*, 6, pp. 263-85.
24. EVANS, J. B., BARSTON, J. L., & POLLARD, P. (1983), ‘On the Conflict Between Logic and Belief in Syllogistic Reasoning’, *Memory & cognition,* 11, 3, pp. 295-306.
25. EVANS, J.B., & OVER, D. E. (1996), ‘*Rationality and Reasoning’,* Hove, England, Psychology Press.
26. EVANS, J.B., & STANOVICH, K.E. (2013), ‘Dual-Process Theories of Higher Cognition: Advancing the Debate’, *Perspectives on Psychological* Science, 8, p. 223.
27. GLANZMAN, D. L. (1995), ‘The Cellular Basis of Classical Conditioning in *Aplysia Californica* – It’s Less Simple Than you Think’, *Trends in Neuroscience* 18, pp. 30-6.
28. GOULD, S. J. (1992), *‘Bully for Brontosaurus: Reflections in Natural History’*, WW Norton & Company.
29. HAWKINS, R. D. (1997), ‘*LTP and Learning: Let’s Stay Together.* *Commentary on Shors, T. J., Matzel, L.D. (1997)*.
30. HEBB, D. O. (1949), *The Organization of Behavior,* New York, John Wiley.
31. HOLENDER, D. (1986), ‘Semantic Activation Without Conscious Identification in Dichotic Listening, Parafoveal Vision, and Visual Masking: A Survey and Appraisal, *Behavioral and Brain Sciences*, Vol. 9, 01.
32. HUANG, Y. & RAO, R.P.N. (2011), ‘Predictive Coding’, *Wiley Interdisciplinary Reviews: Cognitive Science*, Volume 2, Issue 5, pages 580-93, September/October (2011).
33. K. STENNING & M. VAN LAMBALGEN. (2004), ‘A Little Logic Goes a Long Way: Basing Experiment on Semantic Theory in the Cognitive Science of Conditional Reasoning’, *Cognitive Science*, 28(4), pp. 481-530.
34. KAHNEMAN D., & FREDERICK, S. (2002), ‘Representativeness Revisited: Attribute Substitution in Intuitive Judgement’, in Gilovich, T., Griffin, D. and Kahneman, D. (eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*, New York, Cambridge University Press, pp. 49-81.
35. KAHNEMAN, D. (2011), *Thinking, Fast and Slow*. New York, Farrar, Straus and Giroux.
36. KAHNEMAN, D. & FREDERICK, S. (2005), ‘A model of heuristic judgment’, in Holyoak, K. J. and Morrison, R.G. (eds.), *The Cambridge Handbook of Thinking and Reasoning*. New York, Cambridge University Press, pp.267-93.
37. KAHNEMAN, D. & TVERSKY, A. (1982), ‘On the Study of Statistical Intuitions’, *Cognition*, 11,2, pp.123-41.
38. KENNY, A. (1963), *Action, Emotion and Will*, London, New York, Routledge and Kegan Paul, Humanities Press.
39. KEREN, G. & SCHUL, Y. (2007), ‘Two is not Always Better than One’, A Critical Evaluation of Two-Systems Theories. Commentary Barbey & Sloman, (2007).
40. KLAUER, K., MUSCH, J., & NAUMER, B. (2000), ‘On Belief Bias in Syllogistic Reasoning’. *Psychological Review*, 107, pp. 852-84.
41. KOYANO, K., KUBA, K., & MINOTA, S. (1985), ‘Long-Term Potentiation of Transmitter Release Induced by Repetitive Presynaptic Activities in Bull-Frog Sympathetic Ganglia’, *Journal of Physiology*, 359, pp. 219-33.
42. KRUGLANSKI A.W. & GIGERENZER, G. (2011), Intuitive and Deliberate Judgments are Based on Common Principles, *Psychological Review*, 118, 1, pp. 97-109.
43. LAROCHE, S., DOYERE, V., & BLOCH, V. (1989), ‘Linear Relation Between the Magnitude of Long-Term Potentiation in the Dentate Gyrus and Associative Learning in the Rat. A Demonstration Using Commissural Inhibition and Local Infusion of an N-Methyl-D-Aspartate Antagonist’, *Neuroscience*,28, pp. 375-86.
44. LARSON, J. R., & LYNCH, G. (1985), ‘Long-Term Potentiation in the Lizard Cerebral Cortex’, *Society for Neuroscience Abstracts*,11, p. 527.
45. LEWIS, D., & TEYLER, T. J. (1986), ‘Long-Term Potentiation in the Goldfish Optic Tectum’, *Brain Research*,375, pp. 246-50.
46. LIBET, B., KOBAYASHI, H., & TANAKA, T. (1975), ‘Synaptic Coupling into the Production and Storage of a Neuronal Memory Trace’, *Nature*,258, pp. 155-7.
47. LOMO, T. (1966), ‘Frequency Potentiation of Excitatory Synaptic Activity in the Dentate Area of the Hippocampal Formation’, *Acta Physiologica Scandinavica*, 68, suppl., 277, 128.
48. MARTINEZ, C. DO. V., MARTINEZ J. L., and DERRICK, B. E. (2002), ‘Associative Long-Term Potentiation (LTP) Among Extrinsic Afferents of the Hippocampal CA3 Region in Vivo’, *Brain Research*, 940, pp. 86-94.
49. MCDOWELL, J. (1994), *Mind and World*, Cambridge, MA, Harvard University Press.
50. MERCIER, H. & SPERBER, D. (2011), ‘Why do Humans Reason? Arguments for an Argumentative Theory’, *Behavioral and Brain Sciences*, 34, pp. 57-111.
51. MEYER, D. E. & SCHVANEVELDT, R. W. (1971), ‘Facilitation in Recognizing Pairs of Words: Evidence of a Dependence Between Retrieval Operations’, *Journal of Experimental Psychology*, 90, 2, pp. 227–34.
52. MONSELL, S. & DRIVER, J. (2000), *Control of Cognitive Processes*, Cambridge, MA, MIT Press.
53. NORENZAYAN, A., CHOI, I. & NISBETT, R. E. (2002), ‘Cultural Similarities and Differences in Social Inference: Evidence from Behavioral Predictions and Lay Theories of Behavior’, *Personality and Social Psychology Bulletin*, 28, 1, pp. 109–20.
54. NOUDOOST B., CHANG M. H., STEINMETZ N. A., & MOORE, T. (2010), ‘Top-Down Control of Visual Attention’, *Current Opinion in Neurobiology,* 20, pp. 183-90.
55. OAKSFORD, M., MORRIS, F., GRAINGER, B., & WILLIAMS, J. M. G, (1996), ‘Mood, Reasoning, and Central Executive Processes’, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, pp. 476-992.
56. OSMAN, M. (2004), ‘An Evaluation of Dual Process Theories of Reasoning’, *Psychonomic Bulletin & Review*, 11, pp. 998–1010.
57. OSMAN, M. (2013), ‘A Case Study: Dual-Process Theories of Higher Cognition – Commentary on Evans & Stanovich’ (2013), *Perspectives on Psychological Science*, 8, p. 248.
58. PAVLOV, I. P. (1927), ‘*Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex’*, Translated and Edited by G. V. Anrep, London, Oxford University Press.
59. PERRY, J. (2001), *Knowledge, Possibility, and Consciousness*, Cambridge, MA, MIT Press.
60. POCKETT, S., and FIGUROV, A. (1993), ‘Long-Term Potentiation and Depression of the Ventral Horn of Rat Spinal Cord in Vitro’, *NeuroReport*,4, pp. 97-9.
61. POSNER, M. (1978), *Chronometric Exploration of Mind*, Hillsdale, NJ, Erlbaum.
62. PRINZ, J. (2002), Furnishing the Mind: Concepts and Their Perceptual Basis, Cambridge, MA, MIT Press.
63. PRINZ, J. (2005), ‘Are Emotions Feelings?’, *Journal of Consciousness Studies*, 12, 8-10, pp. 9-25.
64. RACINE, R. J., MILGRAM, M. W., and HAFNER, S. (1983), ‘Long-Term Potentiation in the Rat Limbic Forebrain’, *Brain Research*,260, pp. 217–31.
65. RAO R.P.N., and BALLARD D.H. (1999), ‘Predictive Coding in the Visual Cortex: a Functional Interpretation of Some Extraclassical Receptive-Field Effects, *Nature Neuroscience*. 2, pp. 79-87.
66. ROBERTS, M. J., & NEWTON, E. J. (2001), ‘Inspection Times, the Change Task, and the Rapid-Response Selection Task’, *Quarterly Journal of Experimental Psychology*, 54A, pp. 1031-48.
67. Sá, W. C., WEST, R. F., & STANOVICH, K. E. (1999), ‘The Domain Specificity and Generality of Belief Bias: Searching for a Generalizable Critical Thinking Skill’, *Journal of Educational Psychology*, *91*, pp. 497-510.
68. SCOTT, T. R., and BENNETT, M. R. (1993), ‘The Effect of Ions and Second Messengers on Long-Term Potentiation of Chemical Transmission in Avianciliary Ganglia’, *British Journal of Pharmacology*,110, pp. 461-9.
69. SHORS, T. J., MATZEL, L.D. (1997), ‘Long-term Potentiation: What’s Learning got to do with it?’, *Behavioral and Brain Sciences* 20, pp. 597–655.
70. SLOMAN, S. A. (1996), ‘The Empirical Case for Two Systems of Reasoning’, *Psychological Bulletin*, 119, pp. 3-22.

SOLOMON, R. C. (1984), ‘Getting Angry’, in Shweder R. and LeVine R. (eds.), *Culture Theory,* New York, Cambridge University Press, (ch. 9).

1. SOLOMON, R.C. (1976), *The Passions.* New York, Doubleday.
2. SPIVEY, M. J., & GENG, J. J. (2001), ‘Oculomotor Mechanisms Activated by Imagery and Memory: Eye Movements to Absent Objects’, *Psychological Research/Psychologische Forschung, 65*, 4, pp. 235-41.
3. STALNAKER, R. (1998), ‘What Might Nonconceptual Content Be?’, in Villanueva, E. (ed.), *Concepts*, *Philosophical Issues*, Vol. 9, Atascadero, Ridgeview.
4. STANOVICH, K. E. (1999), ‘Who is Rational?’, *Studies of individual differences in reasoning*, Mahwah, NJ, Lawrence Erlbaum Associates, Inc.
5. STANOVICH, K. E. (2009), ‘Distinguishing the Reflective, Algorithmic, and Autonomous Minds: Is it Time for a Tri-Process theory?’, in Evans, J. B. & Frankish, K. (eds.), *In two Minds: Dual Processes and Beyond*, Oxford, Oxford University Press, pp. 55-88.
6. STANOVICH, K. E. (2011), *Rationality and the Reflective Mind*, New York, NY, Oxford University Press.
7. STANOVICH, K. E., & WEST, R. F. (2000), ‘Individual Differences in Reasoning: Implications for the Rationality Debate’, *Behavioral and Brain Sciences*, 23, pp. 645-726.
8. STRIPLING, J. S., PATNEAU, D. K., and GRAMLICH, C. A. (1988), Selective Long-Term Potentiation in the Pyriform Cortex’, *Brain Research*,441, pp. 281-91.
9. TILLAS, A. (forthcoming 2015a), ‘On the Origins of Concepts’, in Kann, C., Hommen, D., and Osswald, T. (eds.), *Concepts and Categorization*. Paderborn, Mentis.
10. TILLAS, A. (forthcoming 2015b), ‘Internal Supervision & Clustering: A New Lesson from “Old” Findings?’, in Müller V. C. (ed.), *Computing and Philosophy*, Synthese Library, Springer.
11. TILLAS, A. (2014), 'How do Ideas Become General in their Signification?’, in Machery, E., Prinz, J. and Skilters, J. (eds.), *The Baltic International Yearbook of Cognition, Logic and Communication*, Vol. 9, Kansas, New Prairie Press.
12. TILLAS, A. (2010), *Back to Our Senses: An Empiricist on Concept Acquisition*, Doctoral Thesis, University of Bristol, UK
13. THOMPSON, V. & EVANS, J. B. (2012), ‘Belief Bias in Informal Reasoning’, *Thinking and Reasoning,* 18, 3, pp. 278-310.
14. UNCAPHER, M. R., HUTCHINSON, B. J. & WAGNER, A. D. (2011), ‘Dissociable Effects of Top-Down and Bottom-Up Attention During Episodic Encoding’, *The Journal of Neuroscience*, 31, 35, pp. 12613–28.
15. WALTERS, E. T. & BYRNE, J. H. (1985), ‘Long-Term Enhancement Produced by Activity-Dependent Modulation of Aplysia Sensory Neurons’, *Journal of Neuroscience,* 5, pp. 662-72.
16. WASON, P. C. & EVANS, J. B. (1975), ‘Dual Processes in Reasoning?’, *Cognition*, 3, 2, pp. 141-54.
17. WILHITE, B. L., TEYLER, T. J., & HENDRICKS, C. (1986), ‘Functional Relations of the Long-Term Potentiation’, *Brain Research,* 365, pp. 54-60.
18. YANG, T. T., GALLEN, C. C., RAMACHADRAN, V. S., COBB, S., SCHWARTZ, B. J., & BLOOM, F. E. (1990), ‘Noninvasive Detection of Cerebral Plasticity in Adult Human Somatosensory Cortex’, *Neuroreport*, 5, pp. 701-4.

**NOTES**

1. University for the Creative Arts, London, UK (jtrafford2@ucreative.ac.uk). [↑](#endnote-ref-1)
2. University of Dusseldorf, Germany (atillas@phil.uni-duesseldorf.de). [↑](#endnote-ref-2)
3. The suggestion is that T-1 processes do not tie-up working memory, and so, they are the default mode of judgment, see [Kahneman & Frederick (2005)]. [↑](#endnote-ref-3)
4. Though Osman [(2004)] defends a uni-model approach. [↑](#endnote-ref-4)
5. See also [Gould (1992)]. [↑](#endnote-ref-5)
6. Though see [Stenning and van Lambalgen (2004)], as well as the discussion in section 4 of this paper. [↑](#endnote-ref-6)
7. See [Dutilh Novaes (2012), esp. ch. 7] for an extended discussion. [↑](#endnote-ref-7)
8. We owe this suggestion to an anonymous reviewer. [↑](#endnote-ref-8)
9. Deontic tasks or tasks that ask people to reason about known regulations, (“If a person is drinking beer, then the person needs to be over 21 years of age”) are associated with T-1 processes. Indicative tasks – tasks appearing within a context that refers to an arbitrary conditional rule, e.g “if there is a vowel on one side, then there is an even number on the other”. [↑](#endnote-ref-9)
10. Note that the point made here is not a semantic one but rather one concerning the makeup of concepts. [↑](#endnote-ref-10)
11. With regards to widely heterogeneously instantiated concepts, perceptually diverge stimuli are brought together in virtue of a sameness marker, e.g. in the case of danger, a mother’s facial expressions play that role, while language plays a crucial role in the process of acquiring lofty concepts. [↑](#endnote-ref-11)
12. See [Tillas (2014)] for a detailed account of abstraction and how type-concepts are formed even though we only have perceptual experiences with particular instances. See also Barsalou’s [(1999)] perceptual simulation theory where he argues that tokening a concept occurs in virtue of simulating the perceptual experience with its instances. [↑](#endnote-ref-12)
13. LTP has been found in various brain regions such as piriform [Stripling et al. (1988)], entorhinal [Wilhite et al. (1986)], and prefrontal cortices [Laroche et al. (1989)], the septum [Racine et al. (1983)], the autonomic [Libet et al. (1975)] and superior [Brown & McAfee (1982)] cervical ganglia, and the ventral horn of the spinal cord [Pockett & Figurov (1993)]. LTP has also been found in vertebrate such as the goldfish [Lewis & Teyler (1986); Yang *et al.* (1990)], bullfrog [Koyano *et al.* (1985)], bird [Scott & Bennett 1993)], and lizard [Larson & Lynch (1985)]. Also, it has been found that LTP occurs in some invertebrates [Glanzman (1995); Walters & Byrne (1985)] reviewed in [Shors and Matzel (1997)]. See also [Lomo (1966); Bliss and Lomo (1973); Bliss and Gardner-Medwin, (1973); Martinez *et al.* (2002)]; see [Shors and Matzel, (1997)] for concerns about the role of LTP in learning processes. For a reply to Shors and Matzel see amongst others [Hawkins (1997)]. [↑](#endnote-ref-13)
14. E.g. [Prinz, (2002)]. [↑](#endnote-ref-14)
15. The above claim about top-down effects in perception is similar to Elman & McClelland’s [(1986)] ‘TRACE Model’. [↑](#endnote-ref-15)
16. See [Spivey & Geng (2001); Chao, Haxby & Martin (1999); Demarais & Cohen (1997)] for evidence in support of this claim. [↑](#endnote-ref-16)
17. Consider also that neurons in different areas of the visual cortex respond selectively to different stimuli, e.g. bars and edges at preferred orientations (V1), while other neurons to complex shapes and contours (V2 and V4), and visual motion (medial superior temporal area). These response selectivities are often understood in terms of hierarchical predictive coding of natural inputs. For example, [Rao & Ballard (1999)] propose a hierarchical neural network in which top-down feedback connections from higher order visual cortical areas carry predictions of lower-level neural activities, while the bottom-up connections convey the residual errors in prediction. See also [Huang & Rao (2011)] for a review, and [Clark (2013)] for a detailed discussion of related issues. [↑](#endnote-ref-17)
18. It is worth clarifying at this point that we allow that some representational content may be nonconceptual, e.g. representations prior becoming parts of concepts. Note that given that pre-conceptual-representations lack the rich cognitive role properties of concepts, it might be superfluous or even misleading to talk about *content*. We do not further elaborate on this issue here, though see [Mcdowell (1994)] for arguments against this line of thought; see also [Stalnaker (1998)] for arguments for nonconceptual content. [↑](#endnote-ref-18)
19. Endogenous control over a given concept is acquired by associating the set of perceptual representations that comprises the concept in question to a perceptual representation of a word or goal-directed actions over which they do have endogenous control. The claim is that concepts inherit the endogenous control that we have over utterances or goal-directed action etc. In a bit more detail, human agents have the ability to manipulate external objects in relationships of agency towards them; what is argued here is that we can piggyback on that ability to manipulate and direct our own thinking. In this sense, given that human subjects have endogenous control over their production of linguistic items, to the extent that we are able to produce linguistic utterances at will (or silent talking to ourselves). It is this executive control over linguistic utterances that gives us endogenous control over our thoughts. See [Tillas (2010)] for a detailed analysis. [↑](#endnote-ref-19)
20. Certain kinds of mental states, like emotions for instance, allow associations to be formed after only one or in any case very few repetitions. [↑](#endnote-ref-20)
21. See [Meyer & Schvaneveldt (1971)] for similar results at the level of semantic representations and sentences. [↑](#endnote-ref-21)
22. The underlying process at hand could perhaps be further understood in terms of ‘semantic activation’ (e.g. [d'Arcais & Schreuder (1987); Dutilh Novaes (2012)]). It is worth clarifying though that the suggested view differs from semantic activation in that: (a) it has a broader scope to the extent that it explains how sub-activated representations become part of a representational network, while setting the background for conceptual flexibility and change (weighting calibration between concepts in a given network Prinz’s [(2002)] cognitive content); (b) unlike semantic activation, which is often understood lexically (or at least to concern lexical concepts), RSA does not require that and so is capable of dealing with the sub-activation of associated images and goal-directed states involving, for example, sensorimotor and proprioceptive representations; (c) resultantly, RSA is a more fine-grained approach than semantic activation. [↑](#endnote-ref-22)
23. It may appear that we have sneaked-in a presupposition that participants already have the concept of “following” in the strict sense of necessary truth-preservation. However, this would be to mistake a more basic sense in which “following” indicates an inferential move or a move in a dialogue, for the formalised notion. On this point, see [Dutilh Novaes (2012)], and the discussion of logical form, below. [↑](#endnote-ref-23)
24. This claim is in line with [Mercier & Sperber (2011)]. [↑](#endnote-ref-24)
25. It is on cases of this kind that DPTs examine and build their argument. However, it is worth clarifying that there are also other cases, which do not challenge prior world knowledge. For instance, having an invalid syllogism with a believable conclusion. Furthermore, there could also be valid syllogisms with believable premises and an unbelievable conclusion. Admittedly, syllogisms of this kind are hard to come by, given the relation of necessary truth-preservation, but they are not impossible. Finally, they could be valid syllogisms with at least one unbelievable premise and an unbelievable conclusion. It is only in the second case that there will be the issue of integrating the (false) premise. In this sense, it would be interesting to see experiments dealing with these kinds of syllogisms and how DPTs interpret the results. We owe this suggestion to an anonymous reviewer. [↑](#endnote-ref-25)
26. For instance, Awh et al. [(1999)] found that storing a location in WM led to faster detection – for targets presented at that location. Downing’s [(2000)] proposal that the storage of an item in WM leads to automatic orienting towards similar objects in the environment; see also [Cowan (2006)]. [↑](#endnote-ref-26)
27. See also [Kruglanski & Gigerenzer (2011), p. 106)]. [↑](#endnote-ref-27)