2015

On the Role of Probability in Hume's Imagination and Associationism: a Bayesian Response to Fodor's Hume Variations

Jason Krivo Flores
Portland State University

Follow this and additional works at: <https://pdxscholar.library.pdx.edu/honorstheses>

Let us know how access to this document benefits you.

Recommended Citation

Flores, Jason Krivo, "On the Role of Probability in Hume's Imagination and Associationism: a Bayesian Response to Fodor's Hume Variations" (2015). *University Honors Theses*. Paper 204.
<https://doi.org/10.15760/honors.123>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in University Honors Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

On the Role of Probability in Hume's Imagination and Associationism: A Bayesian Response to Fodor's *Hume Variations*

JASON KRIVO FLORES

An undergraduate honors thesis submitted in partial fulfillment of the

requirements for the degree of

Bachelor of Arts

in

Philosophy

Thesis Adviser

Dr. Angela Coventry

Portland State University

2015

ACKNOWLEDGMENTS

First and foremost I would like to thank my thesis advisor Dr. Angela Coventry who first suggested the idea for this project. She said something to the effect of: “Hume scholars have taken it surprisingly easy on Fodor” a comment I found incredibly inspiring. Were it not for her patience, encouragement, and constructive critiques I might not have undertaken this project. I also would like to express my gratitude to Dr. Joshua Fost whose Philosophy of Cognitive Science course introduced me to the myriad wonders of “The Bayesian Brain” and thereby to Andy Clark’s elaboration of the predictive processing paradigm which features largely herein. Though the aforementioned scholars are causally implicated in the existence of this manuscript, any errors are solely my responsibility.

I’d also like to thank Avram Hiller for his support and mentoring, Tom Seppalainen, David Weber, and Larry Bowlden from PSU, as well as my friends and teachers (you know who you are) from the 2013 European Summer School in Logic Language and Information (ESSLLI) for introducing me to many of the ideas explored in this work. Additional thanks to the Ronald E. McNair Scholars program whose summer research fellowship provided me with much needed financial support. Lastly, as always, thanks to my chosen family, extended friend family, and progeny, you are the extension to my intension.

J. Krivo Flores
Portland, 2015

I.

On the Role of Probability in Hume's Imagination and Associationism: A Bayesian Response to Fodor's *Hume Variations*

In his 2003 book *Hume Variations*, Jerry Fodor attempted to modernize David Hume's theory of ideas in an effort to bring the latter into accord with the modern field of cognitive science; a field which Fodor sees the classical Computational Theory of Mind (CTM)¹ as exemplifying. In the work, Fodor contends Hume "was largely right about the architecture of the cognitive mind" and even goes as far as to proclaim Hume's *Treatise of Human Nature* as "the foundational document of cognitive science," a document which in Fodor's view comprised the first explicit attempt to construct an empirical psychology on the basis of a representational theory of mind (p. 134). Despite such praise, however, Fodor finds it puzzling Hume's model of cognitive architecture, unlike CTM, includes references to the imagination or "fancy," a fact Fodor takes to evince a flaw in Hume's model rather than his own, contending: "We don't need imagination because we don't need association. And we don't need association because Turing showed us how to replace it with computation" (Fodor, 2003, p. 115).

According to Fodor, associative relations, unlike the computational relations posited by CTM, are unable to distinguish merely *causal* relations amongst thoughts from *intentional* and *logical* relations between the contents of thoughts, e.g., such as distinguishing between a mind which thinks the thought that *P & Q* from a mind which thinks the thought that *P* and is thereby merely *caused* to think that *Q* (p. 115). Fodor believes that Hume, having grasped the above

¹ See, e.g., Fodor, (1975); Newell, (1980); Haugeland, (1981). Technically, Fodor advocates the tripartite theory which includes the Representational Theory of Mind (RTM), Computational Theory of Mind (CTM), and the Language of Thought hypothesis (LOT). I elaborate all three theories in detail in section II. Presently, when I refer to CTM or the Computational Theory of Mind I am including all three theories unless otherwise specified.

shortcoming, invoked the *extra*-associative operations of the imagination as an *ad hoc* solution. Thus Fodor, like much of ‘classical’ cognitive science (e.g., Newell, 1980), equates rational systematic thought as thought, which conforms to the rules of logic—a formal mathematical theory whose domain concerns inferences between propositions which hold with certainty (Fodor, 1975; cf. Oaksford and Chater, 2009). Contrary to such a view, however, probabilistic *Bayesian* approaches attribute the rationality and systematicity of human thought to our ability to adaptively reason under conditions of *uncertainty*—by updating beliefs in response to new evidence (e.g., Oaksford and Chater, 2009; Kruschke, 2008). And while many have noted a recent “probabilistic turn” in cognitive science (e.g., Oaksford and Chater, 2008; Clark, 2012, 2013, 2015; Goodman, et al., 2014), I contend that none, thus far, have fully appreciated just how much such a turn bespeaks a return to Hume, and towards a more sophisticated associationism (Kruschke, 2008, Ryder & Favorov, 2001). Fodor, for example, claims “Hume isn’t really an associationist when push comes to shove, but the price he pays for not being one is that the architecture of his theory rests on the imagination, and the imagination is a something-I-know-not-what” (2003, p. 115).

Pace Fodor, in this paper I will argue that Hume’s basing of his cognitive architecture upon the imagination does not evince a retreat from associationism, but rather shows that Hume’s associationism is of a more sophisticated form than previously acknowledged, I will argue that in founding his associationist architecture upon the imagination Hume portended contemporary *Bayesian* associationist architectures (e.g., Kruschke, 2008; Clark, 2014). For, whereas traditional associationist architectures represent knowledge by simple connection weights (e.g., between the nodes of a neural network) Bayesian associative models represent knowledge as *probability distributions* consisting of graded degrees of belief over a range of

competing hypotheses (Kruschke, 2008). I will argue that this ‘new’ Bayesian associationism (Ryder & Favorov 2001; Kruschke, 2008, Clark 2014, p. 235), is not in fact new, but amounts to a formalization of Hume’s augmentation of traditional associationism by the extra-associative liberty of the imagination. A theoretical move which Fodor finds objectionable, alleging: “all Hume’s theory tells us about the imagination, narrowly construed, is that it accords with the laws of association a lot of the time. *This doesn’t, however, tell us what happens the rest of the time*” (2003, p. 118; emphasis mine). I propose that the answer to Fodor’s is that the essence of the “liberty” contributed by the imagination to Hume’s associationism is best described as *randomness, noise, or stochasticity*—in other words *probability*.

If my analysis is on track, Fodor has misconstrued the Humean mind. I will endeavor to show this is indeed the case by arguing in a key passage cited by Fodor (i.e., THN 1.3.12.22) that Hume is not concerned with addressing the logical issue of scope relations, as Fodor alleges, but with the *psychological* issue of belief revision (cf. Fodor, 2003, p. 114–115). I will argue that Fodor’s misconstrual results from his attempting to interpret Hume’s fundamentally graded² and probabilistic cognitive architecture in terms of the crisp formal logic required by his commitment to the Computational Theory of Mind (CTM). Thus, my engagements with Fodor in the present paper are mostly negative. However, once I have argued my case for the probabilistic nature of the Humean mind, I will conclude by suggesting that future work employ the methods of Bayesian cognitive science to show how the Humean might advance positive solutions to Fodor’s remaining objections.

² By “graded” I here and throughout this paper mean to denote representations that come in degrees and are thus scalable rather than atomic. An example is how probabilities are represented in Bayesian statistics by subjective *degrees of belief*, which are assigned a (real numbered) value between 0 and 1, where 1 denotes an outcome which is certain to obtain, and 0 an outcome which is impossible with all other values encompassing the continuum between certainty and impossibility (or alternately truth and falsity). Atomic (i.e., *ungraded*) representations in contrast, only capture the extremes of true and false, i.e., 0 or 1.

Motivating the Discussion: Hume, Fodor, and the Future of Cognitive Science

Before I begin I should take a moment to situate the current project in the wider debate and argue for the importance of addressing the specific concerns contained herein. First, though several authors have been critical of Fodor's interpretation and attempted appropriation of Hume (e.g., Biro, 2005; Carruthers, 2005; Sarnecki, 2005; Slezak, 2004), none thus far have attempted a rigorous defense of the Humean imagination and associationism against Fodor's specific charges addressed here nor have any, to my knowledge, offered an empirically plausible alternative modernization of Hume to that proposed by Fodor in *Hume Variations*. Quite to the contrary in fact, as John Biro, though critical of Fodor's reading of Hume, has nonetheless argued that Fodor's RTM is the contemporary model of mind Hume would find most "congenial" (Biro, 1991, p. 258). As such, Biro's construal is starkly at odds with the account I offer here and elsewhere (e.g., Flores, 2015).

Furthermore, while other philosophers have noted correspondences between various aspects of Hume's project to the contemporary discipline of cognitive science (e.g., Biro, 1991, 1993; Bower, 1994; Millican, 1998) none, as yet, have drawn detailed parallels between Hume's model of cognitive architecture and contemporary probabilistic Bayesian models (though Hohwy, 2013, does mention Hume briefly). Most germane to the present work, neither have any theorists to date attempted a probabilistic (specifically Bayesian) elucidation of Hume's accounts of mental representation and associationism and as I do here.

Thus, while Fodor and I both believe Hume's insights may be fruitfully incorporated into contemporary cognitive science, we differ in our opinion as to *which* interpretation of cognitive science best fits Hume. I think probabilistic models (e.g., Clark, 2013, 2015; Kruschke, 2008; Oaksford and Chater, 2009; Goodman, et al. 2014) are most indicative of Hume, Fodor thinks

logical models such the classical Computational and Representational Theories of Mind. It could thus be argued that the present engagement with Hume amounts to little more than a proxy war between competing contemporary models of cognitive architecture—or worse, reduces to a superfluous exercise in Whig history. It is thus a fair question then to ask: why bother bringing Hume into the fray at all? It would be less circuitous simply to debate the merits of the respective theories directly. To the above I offer the following rejoinders.

First, the most direct route is not necessarily the most illuminating. I argue that comparing competing theories of cognitive architecture against the familiar backdrop of Hume brings their contrasts and merits into sharper relief whilst adding historical context. For example, I contend the present engagement with Hume and Fodor highlights the debate as to whether human cognition is best construed as fundamentally precise and logical or as inherently uncertain and probabilistic. These two perspectives happen correspond to Hume's mutually exclusive division of the cognitive mind into operations concerning "relations of ideas," those which yield apodictic or demonstratively certain conclusions (i.e., such as in mathematics, formal logic, and geometry), and those applying to "matters of fact," i.e., the nondemonstrative, abductive, and inductive reasonings of everyday life whose conclusions are "merely probable" (EHU 4.1-2; THN 1.3.12.23).³ Fodor's view exemplifies a model of mind based on the former, Hume and Bayesian cognitive scientists the latter.

Lastly, I argue the outcome of the current discussion also bears upon more scholarly considerations, for though Hume continues to be a highly influential and well studied figure whose ideas remain the source of lively debate, his account of cognitive psychology is widely

³ All citations to Hume's *An Enquiry Concerning Human Understanding* are abbreviated 'EHU' with reference to section and paragraph supplied by the Beauchamp edition (1998). Citations to Hume's *A Treatise of Human Nature* are abbreviated 'THN' with reference to the book, part, section and paragraph supplied by the Norton & Norton edition (2000). References to Hume's *An Enquiry Concerning the Principles of Morals* are abbreviated 'EPM' with reference to section and paragraph supplied by the Beauchamp edition (1998).

assumed to be implausible at best and thoroughly wrong-headed at worst. For instance, no less than Barry Stroud has asserted:

If we insist on locating Hume's importance in his naturalistic science of man, it might easily seem that the importance fades [...] If his contributions are to be judged as part of the empirical science of man [...] then his 'results' will appear ludicrously inadequate, and there will be no reason to take him seriously. [...] It can easily be felt that the growth in sophistication complexity, and rigour of the social sciences has left Hume forever behind (Stroud, 1977, p. 223).

If the contemporary Bayesian probabilistic interpretation of Hume, which I defend in this paper, is tenable, Stroud, Fodor, and others like them may soon be required to reconsider their opinions. Should the Bayesian approach to cognitive science prove correct, I contend, contrary to Stroud, that not only does Hume still have much to offer the current (cognitive) sciences of man, but most of what he had to say about the mind may have been correct. I begin with an overview of Fodor's logical and computational account of the mind in order to contrast it with Hume's probabilistic vision.

II.

The Mind According to Fodor

Fodor has described his overall philosophical project as a principled defense of 'folk psychology' also known as *propositional attitude*⁴ psychology, or simply, *belief/desire* psychology (Fodor, 1983, 1987). The term "folk psychology," quite simply, refers to our commonsense intuitions that the best way to predict and account for people's behavior is to understand what they *believe* and what they *desire*. Fodor's model of cognitive architecture consists of three primary components: The *Representational Theory of Mind* (RTM), the

⁴ Propositional attitudes are the thoughts described by such sentence forms as "*S believes that P*," "*S hopes that P*," "*S desires that P*," etc., where "*S*" refers to the subject of the attitude, "*P*" is any sentence, and "*that P*" refers to the proposition that is the object of the attitude. If we let "*A*" stand for such attitude verbs as "*believe*," "*desire*," "*intend*," etc., then the propositional attitude statements all have the form: *S A's that P* (Aydede, 2010).

Computational Theory of Mind (CTM), and the *Language of Thought Hypothesis* (LOT). RTM is the claim that mental processes consist of causal sequences of semantically evaluable symbolic particulars, i.e., mental representations which possess a *content* which can be regarded as either true or false (Fodor, 1987 p. 17). CTM adds the additional thesis that propositional attitudes consist of *computational relations* to the aforementioned mental representations (Clark, 2014 p. 48).

In order for such a model to be true of human beings, however, requires the existence of an inner system of symbolic mental representations with a *compositional* structure, i.e., that the meaning of complex propositions is a deterministic function of their constituent atomic symbols and their arrangement. Fodor asserts, this is needed to explain how thought can be both systematic and productive (Fodor, 2003 p. 135). He calls this hypothetical code, *Mentalese* or the *Language of Thought* (LOT), which he postulates to be an innate human cognitive endowment. Fodor believes Turing's theory of computation in conjunction with RTM and the LOT hypotheses demonstrates how our everyday notions of beliefs and desires can be predictively potent by showing how the semantic and causal properties of symbols can be captured via purely *syntactic* relationships (Fodor, 1987 p. 18). Fodor has summarized the conjunction of the above view as follows:

What makes mind's rational is their ability to perform computations on thoughts, where thoughts, like sentences, are assumed to be syntactically structured, and where "computation" means formal operations in the manner of Turing (Fodor, 1998, p. 205).

Fodor's account of cognition thus combines Alan Turing's discovery that the causal properties of formal systems are determined solely by their syntax with results from the formal logical discipline of *proof theory* which have demonstrated that *semantic* relations, such as truth conditions, may be perfectly mimicked by (i.e., are *isomorphic* to) purely syntactic relations (Fodor, 1987 p. 19). Formal models as outlined above thus demonstrate how semantics may be

yoked to syntax, and when combined with Von Neumann's architecture for a 'computing machine' establishes a means for connecting syntax to causation along with a physical implementation. Put simply, computers *just are* environments in which the syntax of a symbol token determines its causal role in such a way as to respect an assigned semantic content (Fodor, 1987 p. 19). If computational machines are appropriately set up, they will transform one symbol into another only if its syntactic operations mirror the requisite semantic relations—ideally, the same relationship which the premises of a valid logical argument bear to its conclusion, which is why such formal operations are said to be *truth-preserving* (Fodor, 1987 p. 19). The foregoing ideas are well captured by John Haugeland's oft quoted remark "If you take care of the syntax, the semantics will take care of itself" (Haugeland, 1981).

A concrete example of the phenomenon described by Haugeland above is how the purely mechanical (i.e., syntactic) operations of a pocket calculator are deliberately set up to precisely mirror meaningful arithmetical relationships. Fodor's CTM and 'classical' cognitive science is thus predicated on the assumption that our minds function in much the same manner as a calculator, but on a much grander scale (Fodor, 1987, p. 19). Hence, as Fodor himself has stated: 'classical' cognitive science may be rightfully described as "an extended attempt to apply the methods of proof theory to the modeling of thought" (Fodor and Pylyshyn 1988, pp. 29–30) .

It is not hard to appreciate the appeal of the above Turing machine conception of the mind. If such a view is correct then accounting for mental causation and the productivity, compositionality, and systematicity of thought can be explained as a straightforward matter of deductive inference—as Turing machines are logical symbol manipulators *par excellence* (2003, p. 135). Inferring Q from, P & $(P \rightarrow Q)$ for such a device is a simple matter of *modus ponens*: a simple derivation easily physically implemented with basic logic-gates. Fodor argues, therefore,

that we should abandon Hume's inchoate imagination-based associative cognitive architecture in favor of a better understood mental architecture based on the digital computer (Fodor, 2003, p. 131).

In Fodor's view, the above RTM/CTM/LOT model of cognition has obviated the theoretical pressures which motivated Hume to invoke a dedicated faculty of the imagination. Fodor contends since syntactic computation *a la Turing* is capable of explaining everything the imagination can about how the mind's prior experiences affect its current operations, a separate faculty of imagination becomes otiose (2003, p. 130). He thus charges, "We don't need imagination because we don't need association. And we don't need association because Turing showed us how to replace it with computation" (p. 115).

Problems with the Computational Theory of Mind

Contrary to Fodor, however, I believe abandoning the imagination in favor of CTM is the wrong move, for as Fodor himself is increasingly beginning to note (e.g., Fodor, 2008), the Turing machine model of the mind is increasingly plagued by conceptual problems suggesting, at best, it is only likely to account for a limited portion of cognitive capacities. While Turing machines excel at operations based on deduction (something difficult for humans) they are woefully inept when it comes to *abduction*: inference to the best (or likely) explanation based on noisy or incomplete data, something human beings naturally excel at. Furthermore, the problem of *locality vs. globality* captures the apparent fact that the majority of adaptive human intelligence—such as belief formation and hypothesis testing—consists of *non-deductive* rationality dependent on *global* features of the entire cognitive system, which as such do not appear to be based on syntactic operations in the Turing sense. While Hume was unable to

explain the mechanisms behind the mind's ability to reason using global inferences, it is clear that he was aware that such inferences were commonplace. Instructively Hume attributed global inferences to the imagination:

Nothing is more admirable, than the readiness, with which the imagination suggests its ideas, and presents them at the very instant, in which they become necessary or useful. The fancy runs from one end of the universe to the other in collecting those ideas, which belong to any subject. *One would think the whole intellectual world of ideas was at once subjected to our view, and that we did nothing but pick out such as were most proper for our purpose* (THN 1.1.7.15; italics mine).

Contrast the above with the Turing machine notion of computation upon which Fodor's CTM hinges, which consists of 'local' operations upon mental representations in virtue of their logical form, or more simply put, upon the local relations between a complex symbol and its *constituents*. An example: the inference from $(P \ \& \ Q)$ to P , depends purely on the relation between the complex symbol $(P \ \& \ Q)$ and its constituent P . In Fodor's CTM this is interpreted as involving operations over complex mental representations, *qua* sentences in Mentalese encoding the requisite logical forms. However, properties like simplicity, relevance, or explanatory optimality, which appear to guide and constrain human nondemonstrative cognition don't appear to be at all syntactic in the above sense, but rather seem to depend upon *global* properties of the entire cognitive system, or as Fodor puts it:

[W]hat's the best of the available explanations depends on what alternative explanations are available; and, by definition, the presence or absence of alternatives to a hypothesis isn't a local property of that hypothesis (2008, p. 121).

Simply stated: nothing about a belief *qua* complex symbol such as $(P \ \& \ Q)$ and its constituents determines how much endorsing that belief should affect one's *total* set of cognitive commitments (Fodor, 2008, p. 121). Or as Fodor rightly puts it "the truth of $P \ \& \ Q$ is (roughly) a function of the truth of P and the truth of Q ; but the *consistency* of $P \ \& \ Q$ " isn't merely as

function of the consistency of P and the consistency of Q , but rather depends on the *global* consistency of one's *total* set of beliefs (Fodor, 2008, p. 121; footnote 33).

Despite such misgivings however, Fodor still cleaves to CTM arguing that while the leading alternative associationist cognitive architectures (i.e., connectionist machines) fare well at *induction* (i.e., generalizing from experience), such models do no better at abduction. *Pace* Fodor, classical connectionism is not the only viable alternative to CTM, though I do not have the space to adequately discuss all of the relevant details here, probabilistic Bayesian models of cognitive architecture are beginning to tackle the problems of abduction and globality which have stymied logic based approaches such as CTM (for an overview see: Shi & Griffiths, 2009; Clark, 2014, p. 49, 2002; Bonawitz, et al., 2011). Thus as even Fodor is becoming increasingly skeptical that the RTM/CTM/LOT model can ever claim to be a general account of cognition (Fodor, 2008, p. 101), many cognitive scientists are beginning to turn away from logic, the calculus of certainty, towards probability: the calculus of *uncertainty*. I argue that this probabilistic turn in cognitive science comprises a return to Hume. To make this case I will begin by illustrating graded and probabilistic nature of Hume's cognitive architecture and show how the tools of Bayesian cognitive modeling provide an apt means of formalizing his prescient but informal insights.

III.

Hume in the Light of Bayesian Cognitive Science

“[A]ll knowledge resolves itself into probability, and becomes at last of the same nature with that evidence, which we employ in common life” —(THN 1.4.1.4).

Many scholars have noted a similarity between ideas originating in the work of Thomas Bayes (1703) and those of Hume (e.g., Poitras, 2013; Gower, 1990; Owen, 1987). However, the bulk of the literature on Hume and Bayesian probability, which remains controversial, concerns not Hume's theory of ideas or cognitive architecture, but instead is almost entirely concerned with Hume's arguments against human testimony as evidence for the existence of miracles and similar critiques of religion (EHU 10.1; Dawid & Gillies, 1989; Gower, 1990; Owen, 1987; W. Salmon, 1978; Poitras, 2013). A few exceptions are Raynor (1980), who presented evidence that Hume was at least cursorily aware of Bayes' theorem by 1767 (long after Hume composed the works which I here examine), and Gower (1990) who argued that Hume's probability of causes is non-Bayesian in character. However, I consider Gower's arguments to have been soundly refuted by Mura (1998) who has shown Hume's ideas on probability, suitably formalized, comprise a sound system of inductive logic, which is resolutely Bayesian. Space considerations do not permit me to engage the above topics, and moreover, my purpose is largely orthogonal to such debates, being to demonstrate the consilience that results from a Bayesian elaboration of Hume's account of belief, associationism, and the imagination.

My current use of the adjective "Bayesian" in describing Hume should thus not be interpreted too broadly, when using the term I will attempt to be clear as to what property I intend it to denote. Presently, I mean no more than to signify that Hume construed probability as subjective degrees of belief, a conclusion reached by many beside myself (e.g., Mura, 1998; Pearl, 1988 p. 29; Hájek & Hall, 2002). Additionally, however, I will argue that for Hume desires, hopes, and other propositional attitudes (e.g., THN 2.3.9.6) were likewise similarly graded. In fact, it is my strong contention that Hume conceived of mental representations *generally* as being both graded and probabilistic (THN 1.4.1.4-5). I suggest this conclusion,

which some might find controversial, in fact requires little or no exegesis. Hume explicitly and iteratively states that beliefs and desires (e.g., THN 1.3.12.17, and 1.3.12.24, respectively) come in degrees, and that impressions and ideas form a graded continuum (EHU 2.3; THN 2.1.11.7), and that the degree of assent which we grant to any of the above continua is distinguished solely in terms of their relative degrees of *force and vivacity* (THN 1.3.10.3). However, if impressions and ideas are graded and the force and vivacity by which the two are distinguished is graded as well, then *a fortiori*—all mental representations are graded, because for Hume there is nothing other than impressions and ideas.

I will endeavor to show that a Bayesian formalization of Hume's associationist architecture straightforwardly follows from the above premise, a premise for which I will provide a more detailed argument for in a moment. Before I begin however, it is important I briefly take a moment to forestall some tempting misunderstandings as to the nature of my endeavor. First, at no point am I suggesting that there was any direct influence of the work of the Reverend Thomas Bayes on Hume, and—it should be superfluous to emphasize—neither am I implying that Hume thought of what he was doing as “Bayesian” or as having anything to do with Bayes himself or the latter's ideas on probability. Also, the term “Bayesian” refers to the modern field of Bayesian statistical inference as applied to cognitive science, none of which was developed by Thomas Bayes who merely derived the primary theorem which bares his name. Thus, there is in fact very little of the Reverend Bayes' ideas in contemporary Bayesianism.

Second, throughout my exposition I will draw together many discontinuous examples from Hume's writings, a strategy which may strike some as procrustean or gerrymandered. However, I argue that while many of my quotations are indeed pulled from their original contexts, I argue that Hume's view of the mind was unified, and proceeded along the assumption

“It is probable, that one operation and principle of the mind depends on another; which, again, may be resolved into one more general and universal” and he applied the same overarching paradigm to all of his inquiries (EHU 1.15). It is this larger paradigm which I here aim to capture, thus the model of cognitive science here developed should be understood as “Humean” rather than strictly Hume’s. Caveats aside, however, it *is* my claim that the view of the mind I will attempt to formalize and elucidate is indeed Hume’s, beginning with his view of the mind as inherently probabilistic.

Hume’s Subjective Interpretation of Probability

Why I am attributing a specifically *Bayesian* model of cognitive architecture to Hume rather than a probabilistic model more generally? My reason for this is that the particular version of probabilistic cognitive science, which I believe to be most indicative of Hume’s own approach to cognitive psychology, centers upon a specific interpretation of probability and not merely on the mathematics of probability itself. The *Bayesian* cognitive science to which I subscribe approaches probability theory not as a formalism for capturing objective facts in the world (e.g., such as outcomes of tosses of dice), but as an optimal calculus for rationally updating one’s *beliefs* about such events (Oaksford & Chater, 2009). Such a view is known as a *subjectivist* interpretation of probability.

Bruno de Finetti, one of the founders of modern Bayesian statistics, began his treatise *The Theory of Probability* (1974) with the bold statement: “Probability does not exist” (p. x). However, de Finetti was not the first to deny the existence of probability as an objective feature of the world, and was preceded in his view by others, notably Hume, who asserted:

“Though there be no such thing as Chance in the world; our ignorance of the real cause of any event has the same influence on the understanding, and begets a like species of belief or opinion” (EHU 6.1).

What de Finetti and Hume meant was that probability does not exist *objectively*— independent from human psychology—but only *subjectively*, as a measure of an agent’s *degree of belief* in the occurrence or nonoccurrence of a particular event (Nau, 2001, p. 90). Or as Hume put the matter: “what the vulgar call chance is nothing but a secret and concealed cause” (THN 1.3.12.1). I suggest an illuminating way of characterizing Hume’s particular subjectivist interpretation of probability is as the view that beliefs are *graded* in order to account for *uncertainty* (cf. Goodman, et al, 2014).

In Hume’s words: “uncertainty is derived from a concealed contrariety of causes” (THN 1.3.12.21). Thus for Hume like contemporary Bayesians, probability, *qua* degrees of belief, is a subjective measure of our ignorance and uncertainty regarding the hidden causes of our observations. I take the following passages as sufficient to establish this conclusion:

There is certainly a probability, which arises from a superiority of chances on any side; and according as this superiority increases, and surpasses the opposite chances, the probability receives a proportionable increase, and begets still a higher *degree of belief* or assent to that side, in which we discover the superiority (EHU 6.2; italics mine).

And in the Treatise:

The belief, which attends the probability, is a compounded effect, and is formed by the concurrence of the several effects, which proceed from each part of the probability. Since therefore each part of the probability contributes to the production of the belief, each part of the possibility must have the same influence on the opposite side; the nature of these parts being entirely the same. The contrary belief, attending the possibility, implies a view of a certain object, as well as the probability does an opposite view. In this particular both these *degrees of belief* are alike. The only manner then, in which the superior number of similar component parts in the one can exert its influence, and prevail above the inferior in the other, is by producing a stronger and more lively view of its object (THN 1.3.12.17; italics mine).

I here contend that we may perceive a defining difference between Hume and Fodor's respective theories of mind: for Fodor the mental representations which comprise the contents of beliefs are atomic, language-like, and propositional, whereas for Hume propositional attitudes are inherently *probabilistic* and *graded*. In fact, as I argued at the beginning of this section, for Hume *all* mental representations and knowledge states are graded and probabilistic, thus inherently encode uncertainty (THN 1.4.1.4-5).

Evidence that Hume's Representational Theory of Mind is Inherently Probabilistic.

Consider Hume's graded distinction between impressions and ideas. Recall, that according to Hume "All the perceptions of the human mind resolve themselves into two distinct kinds," which he denominates *impressions* and *ideas* distinguished solely by their differing degrees of force and vivacity (THN 1.1.1.1). Impressions are "the more forceful perceptions" which occur, such as when information is registered directly from the senses, whereas ideas are: "the faint images of sensory impressions in thinking and reasoning" (THN 1.1.1). Thus for Hume, both beliefs and impressions are graded and distinguished from suppositions and ideas, respectively, solely by their greater degrees of *force and vivacity*, which I discuss in a moment. In addition to the foregoing distinctions, however, I argue Hume also explicitly advanced a graded theory of *desires*:

I say, it is evident, that the passion, properly speaking, *is not a simple emotion, but a compounded one, of a great number of weaker passions*, derived from a view of each part of the object. For otherwise it were impossible the passion should increase by the increase of these parts. Thus a man, who desires a thousand pound, has in reality *a thousand or more desires which uniting together*, seem to make only one passion (THN 1.3.12.24; emphasis mine).

Furthermore, I contend that Hume appears to hold a graded view of propositional attitudes generally, and applies the above probabilistic framework to passions such as hope, grief, and fear. He writes: "When good is certain or *probable*, it produces joy [...] When either good or evil

is uncertain, it gives rise to *fear* or *hope*, according to the *degrees of uncertainty* on the one side or the other” (THN 2.3.9.6; italics mine). Thus it is evident that for Hume the mental representations that comprise the contents of propositional attitudes are graded in order to account for probabilistic degrees of belief and uncertainty. In fact, a better gloss on Hume’s account might be to describe intentional states such as beliefs, desires, hopes, and fears as *probabilistic attitudes* (Aydede, 2005). I thus allege Hume not only posited degrees of belief, but also degrees of *grief*, and quantified both in terms of their relative uncertainty.

A more compressed argument for my above contention that Hume developed a probabilistic approach to propositional attitude psychology goes as follows: Hume stated: “By knowledge, I mean the *assurance* arising from the comparison of ideas” and “By *probability*, that evidence, which is still attended with *uncertainty*” (THN 1.3.11.2; italics mine). However, Hume ultimately concludes that because of unavoidable uncertainties “*all* knowledge resolves itself into probability, and becomes at last of the same nature with that evidence, which we employ in common life” (THN 1.4.1.4; emphasis mine). Hence, I argue that the Humean mind is fundamentally a *probabilistic mind*. And as we have seen above, Hume’s view of probability was qualitative and subjectivist—specifically the ‘modern’ Bayesian view of probability as a measure one’s subjective degrees of belief. It is in the foregoing respect, therefore, that I conclude that Hume’s account of mental architecture is resolutely Bayesian. I will now show how the above aspects of Hume’s theory cognitive architecture are easily captured via the mathematical concept of a *probability distribution*, which I further show is able to formalize Hume’s notions of *force* and *vivacity*. I then proceed to consider how such relates to our engagement with Fodor.

Humean Mental Representations as Probability Distributions

The advantage of Hume's probabilistic approach to propositional attitude psychology is whereas classical computational models, such as Fodor's, have no obvious way of explaining an agent's uncertainty; Hume's account represents it inherently. In Bayesian cognitive science such mental representations may be formalized as *probability distributions* (Kruschke, 2008). I will use one of Hume's own examples to illustrate this concept. In the following passage Hume was concerned with the question of how "a superior number of equal chances operates upon the mind, and produces belief or assent" (THN 1.3.11.8):

[W]e shall suppose a person to take a dye, formed after such a manner as that four of its sides are marked with one figure, or one number of spots, and two with another; and to put this dye into the box with an intention of throwing it: It is plain, he must conclude the one figure to be more probable than the other, and give the preference to that which is inscribed on the greatest number of sides. He in a manner *believes*, that this will lie uppermost; though still with hesitation and doubt, in proportion to the number of chances, which are contrary: And according as these contrary chances diminish, and the superiority increases on the other side, his *belief acquires new degrees of stability and assurance* (THN 1.3.11. 9; emphasis mine).

Take the above six-sided die to have a circle inscribed on two sides and a triangle on the remaining four. Say your friend shakes such a box and turns it upside down on the table, note at this point it is fully determined (i.e., by 'hidden causes'), hence an objective fact, which side *in fact* lies uppermost. However, even before we observe the outcome (the *hidden variable*) we can nonetheless report how much we *believe* whether a circle or triangle lies uppermost, we might even bet money on the outcome. Probability, thus construed, measures our *prior* degree of belief—our bet—that a particular side has come up *before* observing the actual outcome. Judgments regarding prior probabilities may be understood as *conjectures* or *hypotheses* based on our assumptions and accumulated working knowledge. Instructively, Hume referred to probability as "reasoning by conjecture" (THN 1.3.11.3).

In Hume's example above, since the same figure (i.e., triangle) is inscribed on two faces and another (i.e., circle) on the remaining four, thus (assuming a fair die) the objective odds of the uppermost figure being a triangle are $\frac{2}{6} = \frac{1}{3}$, and $\frac{4}{6} = \frac{2}{3}$ of being a circle. Since the odds are greater for a circle most people—based on symmetry, knowledge, or intuition—will believe most strongly (and if forced to would bet more money) that a circle is most likely to be on top. Hume put the process as follows:

Four sides are supposed in the present case to have the same figure inscribed on them, and two to have another figure. The impulses of the former are, therefore, superior to those of the latter. But as the events are contrary, and it is impossible both these figures can be turned up; the impulses likewise become contrary, and the inferior destroys the superior, as far as its strength goes. The vivacity of the idea is always proportionable to the degrees of the impulse or tendency to the transition; and belief is the same with the vivacity of the idea, according to the precedent doctrine (THN 1.3.11.13).

Thus for Hume, probability in the above example quantifies one's degree of belief in the likelihood of an as yet unobserved outcome, however, say we are judging not one but several events? Take the familiar example of the *Galton Box* below (fig. 1).

(fig. 1). The Galton Box⁵

⁵ Image credit: <http://www.alanzucconi.com/2015/09/09/understanding-the-gaussian-distribution/>

Now, Imagine your friend has dropped a coin into the Galton box, before even looking at the outcome, you can probably report how much you believe that the coin has landed in each possible bin. If you have experience with such a device you will expect the familiar hump shape to eventually emerge in the bins and predict that any given coin is more likely to fall in the center column at first and then is only later likely to trickle into one of the side-most bins (Goodman & Tenenbaum, 2015). Your *idea* of this shape may thus be considered to represent your belief in a particular expected outcome over many trials of the above experiment, formalized mathematically this shape is known as a *probability distribution*: one of the basic concepts essential to statistical inference. The above concept of a probability distribution can thus be seen to formalize Hume's notion that a belief "is to be considered as a compounded effect, of which each part arises from a proportionable number of chances or experiments" (THN 1.3.12.16). Beliefs, and their contents⁶ in this view then are probability distributions, but of what?

Possible worlds. In the context of Bayesian cognitive science, beliefs are often construed as probability distributions over some unknown state of the world. In more philosophical terms we may say a belief consists of a probability distribution over *possible worlds*, e.g., such as a single compound judgment representing our belief in the proportion of possible worlds in which the die comes up circle to the total number of possible worlds in which the die is rolled. Though one may think the notion of a possible world too modern or recondite for Hume, I argue the

⁶ The propositional attitude of belief on my reading, then consists of an *idea* which has as its *content* (or in modern terms its *intentional object*) an impression, whether an impression of sense, memory, or reflection. Alternately the content may be an *idea* related to a present impression. Hume writes: "I say, then, that belief is nothing but a more vivid, lively, forcible, firm, steady conception of an *object*, than what the imagination alone is ever able to attain... it is evident that belief consists not in the peculiar nature or order of *ideas*, but in the manner of their conception, and in their feeling to the mind (EHU 5.10; italics mine). Elsewhere he states: "belief is nothing but a strong and lively idea derived from a present impression related to it (THN 1.3.8.11). thus a belief is an idea with another idea for its content. I here am arguing that both beliefs and their contents consist of probability distributions, this requires then that all perceptions (impressions and ideas) are probabilistic.

contrary, and suggest that save for the modern formalisms, Hume conceived of beliefs in very much the same manner, consider the following passage:

To every probability there is an opposite possibility. This possibility is composed of parts, that are entirely of the same nature with those of the probability; and consequently have the same influence on the mind and understanding. The belief, which attends the probability, is a compounded effect, and is formed by the concurrence of the several effects, which proceed from each part of the probability. Since therefore each part of the probability contributes to the production of the belief, each part of the possibility must have the same influence ... (THN 1.3.12.17; emphasis mine).

Thus, I contend it is no great leap to attribute to Hume the above view that beliefs are comprised of mental representations consisting of probability distributions over possible worlds. Otherwise how else could he have conceived possibilities as being composed of *parts*, each of which is assigned a corresponding to a degree of belief? Thus for Hume (so I argue), a propositional attitude represents one's degree of *assent* (as a probability distribution) towards a mentally represented set of possible states of affairs (also a probability distribution). I believe this proposed formalization makes precise Hume's notion of a belief as a particular manner of conceiving an idea (EHU 5.10). In the view I have here advances a belief (or other propositional attitude) is a *higher-order* mental phenomena. I will now contend the foregoing idea of a higher-order mental state may be employed to formalize Hume's notion of force and vivacity.

Force and vivacity as verisimilitude: perceiving is believing. There is a fair amount of scholarly controversy about how to interpret Hume's talk of force and vivacity (*cf.* Seppalainen & Coventry, 2012; Dauer, 1999; Waxman, 1993; Wild, 2011). Much of the debate concerns whether to interpret force and vivacity *phenomenologically*, i.e., how it feels or seems or *functionally*, i.e., how it behaves in our minds. While I believe there often is a phenomenological aspect to vivacity (we can usually know via introspection whether we are perceiving or imagining a state of affairs, or believe in x as opposed to merely suppose that x), I favor the

particular functional interpretation offered by Wayne Waxman, whose interpretation I believe naturally accords with the account found in Bayesian cognitive science. Waxman argues Hume's notion force and vivacity is best summarized as *verisimilitude*, i.e., "truthlikeness" (Waxman, 1993). By "verisimilitude" is meant precisely what Hume states below:

[T]hat act of the mind, which *renders realities*, or what is taken for such, more present to us than fictions, causes them to weigh more in the thought, and gives them a superior influence on the passions and imagination (THN 1.3.7.7).

Thus vivacity is that factor which makes the difference to us between what is taken for (or acted upon as) reality or truth, and what is not (note that this is a *psychological* claim and not a metaphysical one). Thus we feel and behave very differently when we sincerely *believe* our house is on fire than when we merely *imagine* or *suppose* it to be aflame. Thus belief, *qua* the communication of vivacity as verisimilitude has the capacity to:

[R]aise up a simple idea to an equality with our impressions, and bestow on it a like influence on the passions. This effect it can only have by making an idea approach an impression in force and vivacity. *For as the different degrees of force make all the original difference betwixt an impression and an idea*, they must of consequence be the source of all the differences in the effects of these perceptions (THN 1.3.10.3; emphasis mine).

As Waxman notes, vivacity as outlined in the quote above, functions as an "independent variable" which can completely transform an idea into an impression (e.g., as in madness or sympathy) or weaken a sensation to the point of actually becoming an idea, as in those situations when an impression is so faint that we cannot distinguish it from an idea (Waxman, 1993, p. 78). I will now show how the graded probabilistic account of cognitive architecture thus elaborated allows us to understand Hume's model of the mind via analogy with contemporary associative architectures which have been converted into Bayesian models.

IV.

Hume's 'Bayesian' Associationism

Kruschke in a recent paper (2008) has suggested that most deterministic models of associative learning can be easily converted into probabilistic Bayesian models with two basic modifications. Kruschke uses these modifications to show how the *Kalman filter* may be understood as a “Bayesification” of the traditional Rescorla–Wagner model of associative learning (Kruschke, 2008, p. 224). I now argue that Kruschke’s interpretation may likewise illuminate Hume’s systematic augmentation of traditional associationism.

The first step in converting a traditional model into a Bayesian model, according to Kruschke, requires making the outcome activations probabilistic rather than deterministic. The simplest solution to this step is to take the original deterministic process and inject noise or randomness at a clearly defined point in the processing chain (Goodman, et al., 2014, p. 4). The result is a then *stochastic* process which unfolds just like original process, except at those points where a random choice is made (*ibid.*).

The second step necessary for converting a traditional associative model to a Bayesian one is ensuring the entire hypothesis space of associative weight combinations is simultaneously represented by a degree of belief for each possible combination (i.e., setting the *prior* distribution). This is for the reason that in Bayesian models agents are assumed to simultaneously entertain multiple candidate hypotheses with differing degrees of plausibility (Kruschke, 2008). Hence, for any moment in time t the agent will believe most strongly in some particular hypothesis, somewhat in others, and even less strongly in the remaining ones (Kruschke, 2008, p. 211). In Kruschke’s example using the Kalman filter, the above step was

accomplished by assuming a *normally distributed prior*. How might the above modifications relate to Hume?

Suppose, for the sake of argument, we now decide to apply the above Bayesian modifications to Hume's particular associationism. If we attempt to do so, however, I contend that what we shall discover is that Hume has *already* implemented the requisite modifications. If this is correct, therefore, I argue that it is both appropriate and no great leap to interpret Hume's associationism as effectively comprising a "Bayesification" of traditional associationism, thus opening the door to understanding Hume's project of a science of human nature in Bayesian terms. To further my argument, allow me to briefly review Hume's associationism.

Hume's Principles of the Association of Ideas: *resemblance*, *contiguity*, and *cause and effect*, purport to capture the regularities by which the imagination recombines simple ideas into complex ideas (THN 1.1.4.1). Hume strongly emphasizes however, that the above principles of association are only "a gentle force which commonly prevails" and we are not to conclude that "without [the principles of association] the mind cannot join two ideas as the fables we meet with in poems and romances put this entirely out of question" (THN 1.1.1.58). To account for the foregoing exception, Hume invokes a "second principle" that of "the liberty of the imagination to transpose and change its ideas" (ibid.). Hume often refers to this more unconstrained aspect of the imagination as "the fancy" whose only constraint is that it cannot form images bearing genuinely contradictory properties—such as of a triangle whose third side is longer than the sum of the two remaining sides (EHU 2.4). Now consider Hume's preceding account of "the fancy" to the first step in the Bayesian modifications discussed above, i.e., the injection of *noise* or *randomness*. I contend that they are in effect the same. In fact, Hume explicitly acknowledged

the presence of chance and randomness in his associationism in several places in both the *Treatise* and *Enquiry*, moreover, Hume again attributes this indeterminism to the imagination:

As all simple ideas may be separated by the imagination, and may be united again in what form it pleases, nothing would be more unaccountable than the operations of that faculty, were it not guided by some universal principles, which render it, in some measure, uniform with itself in all times and places. Were ideas entirely loose and unconnected, *chance alone* would join them (THN 1.1.4.1).

Hume above clearly states that it is not merely “*chance alone*” which joins ideas in the imagination, but also the regular workings of the principles of association, however, this entails that *chance also* is a factor in the workings of the imagination. Additionally, in the *Enquiry* Hume also describes what (I contend) we would today characterize as *noise* or *stochasticity* in the otherwise regular workings of the association of ideas, for example:

Resemblance and proximity always produce a relation of ideas; and where you destroy these ties, however *other accidents may bring two ideas together...* (THN 2.2.8.13).

And:

When a sword is leveled at my breast, does not the idea of wound and pain strike me more strongly, than when a glass of wine is presented to me, *even though by accident* this idea should occur after the appearance of the latter object (EHU 5.20)?

I strongly contend that passages such as the above clearly evince Hume’s supplementation of his otherwise regular (i.e., deterministic) principles association with occasional random variations, noise, or “accidents.” Moreover, I suggest that much if not all of the “liberty of the imagination to transpose and change its ideas” *just is* due to the contributions of random noise (THN 1.1.3.4). In fact, Hume elsewhere states that “liberty and chance are synonymous” (THN), even going as far as to explicitly *define* “liberty” as the very same thing as chance⁷. Lastly, I contend that for Hume to state that the principles of association are only “a gentle force, which commonly prevails,” and thus are not fully deterministic simply *is* to state that workings of association are

⁷ I.e., “According to my definitions, necessity makes an essential part of causation; and consequently *liberty*, by removing necessity, removes also causes, and *is the very same thing with chance*” (THN 2.3.1.18; emphasis mine).

ipso facto probabilistic. As a matter of record, I am not the first scholar to highlight the probabilistic character of Hume's principles of association, in fact John Bricke (1974) made much the same point decades ago. I thus conclude Hume's associationism clearly accords with our first Bayesian criteria above. What then of the second criteria?

The simplest strategy for implementing the second criteria above (setting a prior distribution of beliefs over one's total hypothesis space) would be to adopt what is known as the *principle of indifference* by assigning equal probabilities to every possible outcome resulting in what is (somewhat misleadingly) known as a *non-informative prior*. Here again I argue that Hume again has preceded the contemporary Bayesian view, as he clearly describes how the imagination, when faced with uncertain odds, proceeds by assuming the principle of indifference (cf. Mura, 1998, p. 315-316).

Where nothing limits the chances, *every notion, that the most extravagant fancy can form, is upon a footing of equality*; nor can there be any circumstance to give one the advantage above another. Thus unless we allow, that there are some causes to make the dice fall, and preserve their form in their fall, and lie upon some one of their sides, we can form no calculation concerning the laws of hazard (THN 1.3.11.6; emphasis mine).

Hume even references the principle of indifference directly:

A perfect and total indifference is essential to chance, and one total indifference can never in itself be either superior or inferior to another. This truth is not peculiar to my system, but is acknowledged by every one, that forms calculations concerning chances (THN 1.3.11.5).

Thus, in response to Fodor's claim "Hume has no story at all about what the laws are that govern these nonassociative operations of the imagination" (p. 117). I respond that Hume's associationism is aptly construed as a probabilistic Bayes-enhanced version of traditional associationism. I argue that once this is appreciated it can be seen, contrary to Fodor, that rather than being an *ad hoc* means of addressing *logical* issue of scope relations, the primary purview of Hume's imagination concerns the *psychological* issues of probabilistic *belief revision*. Hence,

contrary to Fodor, if my above arguments are correct then Humean view of the mind is more akin to a slot machine than a Turing machine.

V.

Fodor's Misconstrual of the Humean Imagination

The imagination has been described as the “supreme Humean faculty” (Streminger, 1980, p. 91; Salmon, 1929, p. 336). As E.J. Furlong and Norman Kemp Smith have both noted, the Humean imagination is not merely a faculty of feigning, whimsy, and error, but a rational faculty, which operates in tandem with the communication of “vivacity” in order to produce beliefs—including perceptual beliefs, beliefs of memory, as well as beliefs reached by the understanding; the faculty of the mind which reasons (Furlong, 1961, p. 64). Thus as Furlong has aptly observed: “Where an intellectualist solves problems by recourse to reason or intellect Hume’s universal remedy appears to be imagination” (Furlong, 1961, p. 63). A statement attested to by Hume’s assertion “The memory, senses, and understanding are, therefore, all of them founded on the imagination, or the vivacity of our ideas” (THN 1.4.7.3). However, if the account I have argued for in the preceding sections is correct, the imagination’s role in cognition is best understood in terms of Bayesian inference, and thus belief revision.

I contend therefore, that Fodor’s critique of Hume’s associationism and imagination is largely based on a misconstrual of the following passage (excerpted by Fodor, 2003, p. 115) from Hume’s chapter *Of the Probability of Causes*. I suggest that Fodor has erred by misinterpreting Hume’s fundamentally graded and probabilistic account in terms of the crisp formal logic required by the classical Computational Theory of Mind. Fodor puts forward the following passage as evidence for his claims:

When we transfer contrary experiments to the future, we can only repeat these contrary experiments with their particular proportions; which could not produce assurance in any single event, upon which we reason, unless the fancy melted together all those images that concur, and extracted from them one single idea or image, which is intense and lively in proportion to the number of experiments from which it is derived, and their superiority above their antagonists. Our past experience presents no determinate object; and as our belief, however faint, fixes itself on a determinate object, it is evident that the belief arises not merely from the transference of past to future, but from some operation of the fancy conjoined with it. This may lead us to conceive the manner, in which that faculty enters into all our reasonings (THN 1.3.12.22; emphasis mine).

Fodor, however, interprets the above passage as comprising an “elegant example” of Hume’s apprehension of the need to distinguish merely associative relations among mental states from logical relations among their intentional objects, “But since association provides no way to do so,” says Fodor, the Humean “imagination must come to the rescue” (2003, p. 115).

Pace Fodor, instead of reading Hume above as invoking the imagination to solve a strictly *logical* problem, I argue that we should take the passage at face value—as concerning probabilistic *prediction* and *belief revision*. I argue the passage clearly describes (i.e., “when we transfer the past to the future...”) prediction as a process of belief revision in the imagination (“the belief arises [...] from some operation of the fancy”). A process, which, I contend, is best, described as the synthesis of a probability distribution, i.e., “a single idea,” over graded degrees of belief (THN, 1.3.12.22)? Perhaps, as argued by Oaksford and Chater (2009), human rationality, hence the systematicity and productivity of human thought, is best accounted for not by logic and proof theory, but again by *probability*. I contend that Fodor has surreptitiously inserted cognitivist assumptions into his critique, namely, the assumption that thinking consists of formal logical operations over complex symbols.

I argue the correct reading of the above passage instead requires the recognition that Hume understood cognition as a probabilistic rather than logical process. I contend Hume advocates something quite the opposite of Fodor: that everyday reasoning is expressly *not* the

function of the understanding, i.e., the part of us that reasons logically, but of imagination. To this Fodor is likely to object that Hume's "imagination is a blank check" or that "Hume has no story at all about what the laws are that govern these nonassociative operations of the imagination (2003, p. 117). But as I have already argued, the nonassociative operations of the imagination may be accounted for as the contributions of randomness and uncertainty. But what of the remaining operations of the fancy—those which may account for the systematicity of thought and rational belief revision?

I argue that answer may be found in the very passage cited by Fodor wherein explicitly states that the mechanism whereby the imagination enters into *all* of our reasonings (suitably formalized) is *belief revision* via the synthesis (or sampling) of *posterior predictive distributions*, i.e., the distribution of unobserved observations (i.e., future predictions) conditional on previously observed data. I.e., in Hume's argot: "melting together" images which concur and extracting from them a single idea or image "lively in proportion to the number of experiments from which it is derived" (THN, 1.3.12.22). Hume uses the example of the predicted likelihood of the outcome of a future (unobserved) experiment (i.e., toss of a die), conditional on past assumptions and experiments (observed tosses and the assumption that each side of a 6 sided die is equally likely to come up) (THN 1.3.11.9). I contend this reading is supported by the many places where Hume explicitly describes the predictive, future-oriented, nature of the imagination, e.g., when "the fancy remains fixed, and from the present instant surveys the future and the past" (THN 2.3.7.9), and his assertion: "The fancy anticipates the course of things, and surveys the object in that condition, to which it tends, as well as in that, which is regarded as the present" (THN 2.3.7.9).

In Bayesian cognitive models a posterior predictive distribution may be used to formally represent Hume's notion of "a single idea or image [of a future event], which is intense and lively in proportion (i.e., conditioned upon) the number of experiments from which it is derived" (e.g., THN 1.3.12.12; cf. Lee & Wagenmakers, 2014, p. 45). Hume asserts "our belief, however faint, fixes itself on a determinate object." I claim this "determinate object" of which Hume speaks is, in Bayesian terms, a posterior predictive distribution. Hence the passage, which Fodor interprets as concerned with addressing a logical issue, in fact has Hume offering an apt—yet informal—description of the imagination's role in probabilistic belief revision. I thus conclude that Fodor has misconstrued the Humean imagination, and thus the more appropriate formal analysis of Hume's cognitive architecture requires the methods of (Bayesian) probability rather than formal logic and proof theory.

VI.

Conclusion:

I have thus argued that Hume's model of the mind is inherently probabilistic. I contend that for Hume all mental representations, including impressions, ideas, beliefs, and desires are graded in order to account for uncertainty. I have further alleged that the "liberty" by which the Humean imagination transcends the principles of the association of ideas is the liberty of chance, noise, or randomness. I have also demonstrated that the result of Hume's modification results in a resolutely Bayesian form of associationism. Lastly, I have contended that the above reading suggests that Fodor has misconstrued Hume's cognitive architecture by attempting to interpret what is inherently probabilistic in terms of the crisp formal logic required by his commitment to the Computational Theory of Mind (CTM). If the Humean mind—and indeed the *human* mind—

traffics in probability distributions rather than Fodor's symbols of Mentalese, a proper account of cognition will likely require the tools of probability rather than logic.

It is thus my belief that the formal methods of Bayesian cognitive science discussed in this paper provide the computational tools with which to formalize, unify, and further elucidate many of Hume's original insights, and perhaps, to even continue his goal of constructing a unified cognitive science of human nature. If the account which I have developed in this paper is plausible, Hume's probabilistic approach to propositional attitude psychology was radically ahead of his time, and the current probabilistic turn in cognitive science should be acknowledged, at least in part, to comprise return to Hume.

VII.

References

- Aydede, M. (2005). The language of thought hypothesis. *The Stanford Encyclopedia of Philosophy* (Fall 2010 Edition), Edward N. Zalta (ed.), URL: <http://plato.stanford.edu/archives/fall2010/entries/language-thought/>.
- Bayes, T. (1703). An essay towards solving a problem in the doctrine of chances. *Studies in the History of Statistics and Probability*, 1, 134-153.
- Biro, J. (1991). Hume and cognitive science. In *Historical Foundations of Cognitive Science* (pp. 119-137). Springer Netherlands.
- Biro, J. (1993). Hume's new science of the mind. *The Cambridge Companion to Hume*, 33.
- Bonawitz, E., Denison, S., Chen, A., Gopnik, A., & Griffiths, T. L. (2011, July). A simple sequential algorithm for approximating Bayesian inference. In *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 2463-2468).
- Bower, K. J. (1984). Imagery: From Hume to Cognitive Science. *Canadian Journal of Philosophy*, 14(2), 217-234.

- Bricke, J. (1974). Hume's associationist psychology. *Journal of the History of the Behavioral Sciences*, 10(4), 397-409.
- Brockman, J., Ferguson, A., & Ford, M. (2014). *This explains everything*. New York. Harper Perennial. Print.
- Carruthers, P. (2005). Review: Hume Variations. *Mind*, 114(453), 141-145.
- Clark, Andy. (2014). *Mindware: an Introduction to the philosophy of cognitive Science*. New York: Oxford UP. Print.
- Clark, A. (2015). Embodied prediction. *Open MIND*, 7.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36 (03), 181-204.
- Clark, A. (2012). Dreaming the whole cat: Generative models, predictive processing, and the enactivist conception of perceptual experience. *Mind*, 106.
- Dawid, P., & Gillies, D. (1989). A Bayesian analysis of Hume's argument concerning miracles. *The Philosophical Quarterly*, 57-65.
- Flores, J. K. (2015). Hume in the Light of Bayes: Towards a Unified Cognitive Science of Human Nature. *Res Cogitans*, 6(1), 2-13.
- Fodor, J. A. (1975). *The language of thought*. Harvard University Press.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.
- Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. The MIT Press.
- Fodor, Jerry A. (2003) *Hume Variations*. Oxford: Clarendon.
- Fodor, Jerry. (2004) Having concepts: A brief refutation of the twentieth century. *Mind & Language* 19.1 : 29-47.
- Fodor, J. A.. (2008). *LOT 2: The Language of Thought Revisited*. Oxford University Press.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1), 3-71.
- Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition.
- Hájek, A., & Hall, N. (2002). Induction and probability. *The Blackwell Guide to the*, 149.
- Furlong, E. J. (1961). Imagination in Hume's " Treatise " and " Enquiry concerning the Human Understanding ". *Philosophy*, 62-70.

- Glickman, M. E., & van Dyk, D. A. (2007). Basic Bayesian Methods. In *Topics in Biostatistics* (pp. 319-338). Humana Press.
- Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2014). Concepts in a probabilistic language of thought. *To appear in The Conceptual Mind: New Directions in the Study of Concepts*.
- Goodman and J. B. Tenenbaum (electronic). Probabilistic Models of Cognition. Retrieved, 5/18/2015, from <http://probmods.org>.
- Gower, B. (1990). David Hume and the Probability of Miracles. *Hume studies*, 16(1), 17.
- Haugeland, J. (1981). *Semantic engines: An introduction to mind design*. MIT Press.
- N. D. Goodman and J. B. Tenenbaum (electronic). Probabilistic Models of Cognition. Retrieved, 5/18/2015, from <http://probmods.org>.
- Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
- Horst, Steven, "The Computational Theory of Mind", The Stanford Encyclopedia of Philosophy (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2015/entries/computational-mind/>.
- Hume, D., & Beauchamp, L. (2000). *An Enquiry concerning Human Understanding*. Oxford: Oxford UP.
- Hume, D., & Norton, D.F., and Norton, M.J. (2000). *A Treatise of Human Nature*. Oxford: Oxford UP.
- Kruschke, J. K. (2008). Bayesian approaches to associative learning: From passive to active learning. *Learning & Behavior*, 36(3), 210-226.
- Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Millican, P. (1998). Hume on Reason and Induction. *Hume Studies*, 24(1), 141-159.
- Mura, A. (1998). Hume's inductive logic. *Synthese*, 115(3), 303-331.
- Nau, R. F. (2001). De Finetti was right: probability does not exist. *Theory and Decision*, 51(2-4), 89-124.
- Newell, A. (1980). Physical Symbol Systems. *Cognitive science*, 4(2), 135-183.

- Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences*, 32(01), 69-84.
- Owen, D. (1987). Hume versus Price on miracles and prior probabilities: testimony and the Bayesian calculation. *The Philosophical Quarterly*, 187-202.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.
- Piaget, J. (1976). *The affective unconscious and the cognitive unconscious* (pp. 63-71). Springer Berlin Heidelberg.
- Poitras, G. (2013). Richard Price, miracles and the origins of Bayesian decision theory. *The European Journal of the History of Economic Thought*, 20(1), 29-57.
- Raynor, D. (1980). Hume's knowledge of Bayes's theorem. *Philosophical Studies*, 38(1), 105-106.
- Ryder, D., & Favorov, O. V. (2001). The new associationism: a neural explanation for the predictive powers of cerebral cortex. *Brain and Mind*, 2(2), 161-194.
- Salmon, W. C. (1978). Religion and science: a new look at Hume's Dialogues. *Philosophical Studies*, 33(2), 143-176.
- Sarnecki, J. (2005). Review: Hume Variations, *Dialogue*, 44(04), 809-811.
- Seppalainen, T., & Coventry, A. (2012). Hume's Empiricist Inner Epistemology: A Reassessment of the Copy Principle. *The Continuum Companion to Hume*, 38-56.
- Shi, L., & Griffiths, T. L. (2009). Neural implementation of hierarchical Bayesian inference by importance sampling. In *Advances in neural information processing systems* (pp. 1669-1677).
- Slezak, P. (2004). Fodor's 'Guilty Passions': Representation as Hume's Ideas. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1255-1260).
- Sobel, J. H. (1987). On the evidence of testimony for miracles: a Bayesian interpretation of David Hume's analysis. *The Philosophical Quarterly*, 166-186.
- Streminger, G. (1980). Hume's Theory of Imagination. *Hume Studies*, 6(2), 91-118.
- Stroud, B. (1977). *Hume* (Vol. 5). Psychology Press.

Waxman, W. (1993). Impressions and Ideas: Vivacity as Verisimilitude. *Hume Studies*, 19(1), 75-88.

Wild, M. (2011). Hume on Force and Vivacity. *Logical Analysis & History of Philosophy/Philosophiegeschichte und Logische Analyse*, 1

Zabell, S. L. (1989). The rule of succession. *Erkenntnis*, 31(2-3), 283-321.