THE LOGIC OF LOVE

By

AVIV KEREN

Discussion Paper # 695  (March 2016)

The Federmann Center for
The Study of Rationality

Feldman Building, Edmond J. Safra Campus,
Jerusalem 91904, Israel
PHONE: [972]-2-6584135      FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/
The Logic of Love
Aviv Keren

Love is composed of a single soul inhabiting two bodies. – Aristotle

Abstract
This philosophical work lays the groundwork for a game-theoretic account of (romantic) love, substantiating the folk-psychological conception of love as ‘a unification of souls’. It does so by setting up an appropriate universal framework of cognitive agency, that accommodates such unifications and motivates them. This framework applies the gene’s eye view of evolution to the evolution of cognition, integrating it with a distributed, dynamic theory of selfhood – and the game-theoretic principles of agent-unification that govern these dynamics. The application of this framework to particular biological settings produces love as a theoretical evolutionary prediction (unveiling its rationality). Through this, the connection of the strategic normativity to love’s real-life behavioral and phenomenological expressions is systematically explored.

1 The Federman Center for the Study of Rationality, Hebrew University of Jerusalem.

The author would like to thank Lior Abramson, Robert Aumann, Dan Cohen, Avi Kenan, Uzi Motro, Ilan Nehama, Eyal Winter, and the center and its people in general.

For the full, updated version, please visit https://huji.academia.edu/AvivKeren.
# Contents

Abstract ........................................................................................................................................... 1  

1 Introduction .................................................................................................................................. 3  
   1.2 Related Research – A Brief Overview .................................................................................... 4  
   1.3 The Main Narrative ................................................................................................................. 6  

2 The Cogni-Gene’s Eye View ......................................................................................................... 8  
   2.1 Background: Dawkins .............................................................................................................. 8  
   2.2 Extending the Extended Phenotype ......................................................................................... 13  

3 The Society of Self ....................................................................................................................... 17  
   3.1 Background: Prelude to Philosophy of the Self ...................................................................... 17  
   3.2 Background: Identity and Head of State ................................................................................. 19  
   3.3 Personal Identity Generalized ................................................................................................ 22  

4 Unification and its Dynamics ....................................................................................................... 24  
   4.1 A Union .................................................................................................................................. 25  
   4.2 Uniting ................................................................................................................................... 27  
   4.3 Harmony ................................................................................................................................ 30  

5 The Framework – A Synthesis ..................................................................................................... 34  
   5.1 Background: Evolutionary Psychology .................................................................................... 34  
   5.2 Tying it all up ............................................................................................................................ 36  
   5.3 Background: Mating Behavior and Parental Investment ....................................................... 38  

6 Love ............................................................................................................................................... 40  
   6.1 The Heart of Love ..................................................................................................................... 40  
   6.2 Commitment .............................................................................................................................. 42  
   6.3 Person-Fusion ........................................................................................................................... 44  
   6.4 Other Influencing Parameters ................................................................................................. 48  
   6.5 Beyond Heterosexual and Romantic Love? ........................................................................... 49  
   6.6 Interim Conclusion .................................................................................................................... 52  

Bibliography .................................................................................................................................... 53
1 Introduction

This work is a philosophical, preliminary elucidation; an endeavor to locate the phenomenon of romantic love within the correct conceptual framework in which it belongs: the framework in which it should be understood, could be formalized, and be better explored scientifically. It will manage to give a principled account while sticking mostly to abstract principles of cognitive agency. This will allow us to offer a basic framing and understanding of the phenomenon without getting into the many empirical details and complexities of emotions, the brain’s inner-workings, and the way these are brought about by the genes\(^3\). In a single phrase, this work is about establishing Aristotle’s metaphor.

1.1.1 Establishing the Principal Metaphor

"Love is composed of a single soul inhabiting two bodies". Although the soul is an archaic term, it is very much still with us, as the self: that entity to which "my body" belongs, that mental entity that is the product of the brain. But Aristotle’s metaphor itself is still untenable if the self is taken to be metaphysically primal, indecomposable (into multiple selves)\(^4\), an entity that is – by its very essence – tied to an individual organism. So this work as a philosophically-destructive one, that makes progress by shooting backwards, aims for the liberation from a predominant preconception:

*The Central Dogma:*  One body – one self.

While the body is determined physically, the self involves a conceptualization on our part, concerning *behavior attribution*. In case of deep inconsistencies in a person’s behavior, for example, we may find it more plausible to split that attribution into a few different selves – this is the psychological phenomenon popularly known as "Multiple Personality Disorder". And love, as shall be our main postulate, concerns the opposite direction: the attribution of behavior to a single entity despite that overall behavior originating from different bodies – making the most sense. The fundamental issue thus is: who the players/agents are (Aumann, 2012). And the main prerequisite – an open mind. In particular, one must unequivocally resist a very common conception of love, as an irrational emotion.

1.1.2 Emotions and Rationality

In line with the centrality of the individual in (Western) psychology (both folk and academic), emotions are conceived within a framework of a fixed, well-defined self. The simplistic dichotomy opposing emotions with the rational requires a fixed agent with its well-defined self best-interest, in pursuit of which it can either act correctly or not. And it is with regard to that same agent that higher-level possible explanations for a behavior’s rational origins – a fundamental premise and working-hypothesis of evolutionary psychology – are sought. E.g., fear of heights or snakes as innate, unconscious knowledge. Or the urge for revenge as rationality within a larger-scale, repeated interaction, where reputation is involved.

What about love? Some of its aspects can conceivably figure into an individualistic rational framework. Falling in love 'with the wrong person', for example, could perhaps reflect people’s conscious theories not getting their evolutionary purposes right (e.g. religiously or culturally driven faithfulness to the lawfully-wedded spouse); or innate dispositions that are unadjusted

---

\(^3\) While avoiding the notorious threat of "just-so stories" that is prevalent in evolutionary psychology.

\(^4\) It may certainly have distinct sub-entities for parts, as for it does Aristotle, Freud, and many others.
to modern life ("preferring muscles over brains"). But other aspects... cannot so easily be explained away.

Heartache
Consider John, who has recently been dumped by Jane, unexpectedly. Generally, he is successful and handsome (pardon the illustrative simplicity), and has already had quite a few girlfriends in the past, including three with whom he was even in love, and happy for the most part. Objectively, his situation is quite ok; surely he could be quick to find another with whom he’ll be happy again. Yet he – completely irrationally, it would seem – "insists" on being depressed and dysfunctional, obsessing over the lost relationship and its futureless past, avoiding friends, ditching work, claiming to feel as if there’s no point in going on living, and de-facto behaving as if he means it. Could that be rational?

Frank (1988) combines two elements from game theory to embed emotions within a general rational framework: The strategic value of commitment, of the ability to give up on certain options, and the importance and challenge of being able to signal that commitment credibly. If one can convince in advance that he will be vindictive if harmed (i.e., forgo the option to then be rational about it), he can motivate others not to harm him – all matters of repeated interactions and reputation aside. Emotions are thus rational tools for solving the commitment problem. And love – a particular commitment:

Going through life, one can only sample and estimate the highest quality of mates that is available to him. He has to stop scouting and make a choice at some point – with higher-quality mates certainly still out there. Stepping outside the comfort-zone of self-centered rationality, love is thus the commitment not to simply switch to a better mate when she comes along in the future, as they certainly may. A commitment that (by requirement, by design) can be signaled – credibly so. Such credibility may involve displays of "crazy love", which Frank is thus able to give reasons for, without getting too deep into the actual psychological implementation.

His account certainly takes important first steps towards showing how the infamously-irrational phenomenon of love does not in fact have to contradict rationality, but just the opposite. But can this alone explain heartache's excessive intensity? And more generally, does it actually capture the phenomenon in full, with all its behavioral and experiential implications as familiar to us? With its deep potential impact on our mental existence, its power to change us? Labeling love as another emotion, and simply condemning it as (seemingly) irrational with respect to the individual, frames the problem as one of finding the picture within which that individual is being rational. This, on our view, misses the whole point of with respect to what love is rational (which is not the organism). Developing this view will put us in a position to see the partiality of Frank's suggestion and how it is embedded within a much richer picture.

But much more generally than just for love, even the fact that in so many other cases the identity of the 'correct' agent is clear, is far from trivial, and from a modern evolutionary point of view, actually in need of serious justification. Understanding this background (with which this work will open) will allow us to integrate love properly into a framework of rationality.

1.2 Related Research – A Brief Overview
Most of the works on love throughout the ages, were done through methodological approaches and spelled out in terms that do not come close enough to being scientific (unless one considers psychoanalysis to be a science, obviously), and certainly not mathematizable. This does not mean that they are all irrelevant and bear no connections to our own account here – quite the
opposite. But for the more focused task of the necessary comparisons with existing research and assessment of our own contribution – we take them to be outside the scope here.

Before getting into natural minds and their particulars, there is the field of artificial intelligence – aiming at the construction of artificial, engineered "minds", exploring their general design principles. Common views opposing emotions with rationality may have somewhat deterred the field from paying due attention to emotions (and certainly love). But there are other voices, making room for emotions within a general viable A.I. architecture. One noteworthy work is Wright et al. (1996), which accounts for (some aspects of) love right at the implementation in the underlying information-processing architecture. It is generally compatible with our high-level strategic account.

Modern empirical psychology’s take on love is best represented by Sternberg et al. (2006). The many different works there share a quite descriptive nature (trying to account for the facets of love through a few ‘basic’ ingredients that are just taken as facts, etc.). The more prescriptive evolutionary perspectives at best just posit love as for the benefit of the offspring, without getting into the details we will be going into (the genes’ eye-view and the particular strategic advantage). It is my personal impression that the works there could have a lot to gain from a more theoretical grounding such as ours.

Psychology aside, one could simply focus on the strategic aspects of love as a behavior. That would be the game-theoretic or economics approach. The work closest to ours on this score is Frank’s (1988) (combining a like-minded economics approach with some cognitive findings and evolutionary justifications as well). This work does a great job directing the search for rationality towards squaring with our non-strategic, psychological life-acquaintance with love (contra simplistic materialist views following works like Becker’s (1993)). However, his framework is ultimately one of organism-rationality – which precludes most of our stronger derivations and misses out on the depth of the phenomenon (strategic as well as psychological). So we will just focus on developing our own account in full instead, and point to important differences as they arise.

Bringing the two former fields and perspectives together is evolutionary psychology – within which our work, too, belongs (if anywhere). It is heavily invested in much-related issues: mate selection and mating behavior, parental investment, sexual conflicts and sexual competition etc., always through the lens of strategy-analysis. These issues are paradigmatic, and perhaps where the field excels the most, in terms of its predictive and explanatory power and success. However, these usually analyze and observe particular concrete strategies. The much more abstract topic of love, too, is beginning to receive such exploration (Campbell, et al., 2005; Buss, 2006). Our own approach will integrate deeper themes from game theory and the philosophy of mind than the works and standard modes of reasoning in the field, to handle this far-reaching overall psychological transformations that is love. We will review the evolutionary psychology background and fit our own approach into it in due course.

Like the former fields, philosophy also considers (at least central aspects of) the subject as its own (through its own lens, questions, methodologies). And like psychology, it usually sticks to

---

5 including central ones, such as Marvin Minsky (The Emotion Machine, 2006).
6 Colyvan (2010) approaches love from a purely game-theoretic perspective, taking it for a mechanism for ensuring long-term cooperation. But it does so without Frank’s (1988) systematic approach to commitment mechanisms, and generally amounts to only very brief remarks, being mostly an introduction to such a game-theoretic approach.
ultimately descriptive accounts or at least relies on contingent components of how humans actually are (of the debatable high-level psychological sort, not just their basic biology). One notable historical figure, who could be considered the founding father of our approach, is Schopenhauer (whose "ideas foreshadowed the discovery of evolution, Freud's concepts of the libido and the unconscious mind, and evolutionary psychology in general" [Wikipedia]). Specifically, he views love as driven by the unconscious goal to procreate (independently of the person' conscious goals). And frames the choice of a mate, in such terms, as the best 'completion' for the task. But his account (Schopenhauer) is (understandably) extremely outdated in many central respects. We won’t get into complicated comparisons, and just try to get it right. As for modern philosophy: "Love as Union" is a common type of account\(^7\) (of which our own account is one), but the related issues dealt with are usually different. Specifically, I am not aware of accounts similar to ours, including with respect to: the multi-agent view; the extension of the gene's eye-view (which is considered much more scientifically confirmed or at least grounded than philosophical theories can usually hope to be); tying up to independently-motivated theories of consciousness (or at least the particular one we will be embedding, which suggests further connections and explorations); and the 'game-theoretic' analysis we will be taking on. From the outset, part of the value of this work is in its fresh framing of the issue, borrowing inspiration and insight from and heavily relying on other fields (with their independently-valued intellectual progress). This value would be dragged down by sticking to philosophy’s classical and common framing of the topic. Thus, we will avoid making these necessary connections to the existing philosophical literature here.

1.3 The Main Narrative

We will present a framework that synthesizes theories on two levels: an underlying evolutionary-genetic level, and a psychological one on top. Each level’s own purpose is to make room for the possibility of agent-unification (of the agents of that level). We will then attend explicitly to unification as the fundamental (though unexplored) game-theoretic issue that it is, and illustrate it. A principle of cooperation (a form of symbiosis) will give reason for making such unions.

\[ \text{Figure 1: The Framework – a Synthesis} \]

\(^7\) [http://plato.stanford.edu/entries/love/](http://plato.stanford.edu/entries/love/)
Following some mandatory basics from evolutionary psychology, we will be able to bring all of these together: With the psychological agents already in place, the (standard) argument then is that, evolutionarily, the genes ought to implement the psychological level so as to carry over their own interests, including their interests to unite when appropriate (according to the abstract unification principles put forth). From this, in the context of sexual reproduction, love will result as a prediction.

We will examine the ideal of love, and make sense of it in terms of unity (of cognitive agency). Making sense of it in strategic terms of commitment will uncover the complexities that arise in non-ideal settings, and the influence of various biological and environmental factors. Through these lens we will make sense of many behavioral and phenomenological aspects of love: "Irrational" expressions such as post-breakup depression; the power to change a person; the loved appearing to be "special"; and more.

We will be relying extensively on existing theories (that shall be introduced as the story unfolds). Some are firmly established and accepted, and all originate from independent motivations. Assessing how close our necessarily-speculative story is to being factually true, is a highly non-trivial matter (which we defer to part II). But the underlying issues at stake will be illuminated in any case, we believe. Actually becoming convinced of our approach, takes a bigger philosophical leap than this text can support within its bounds. The prerequisites for that are a thorough acquaintance and acceptance of the background we will be introducing only briefly (particularly the gene’s eye view, and evolutionary psychology and its style of 'reductionism', that people may tend to resist for various reasons external to our topic).

This paper (considered here as Part I) presents the central narrative of our account for the phenomenon. Related approaches and works on which our account does not build directly will be mentioned only briefly. The aim of the paper is to give a fairly complete and convincing but focused account, the conceptual core of our approach. That core interacts with many related issues, shedding new light or offering interesting critiques of related approaches, and suggesting further developments and research. To all of these, a hypothetical Part II would be dedicated (sketched and begun in the full version, at https://huji.academia.edu/AvivKeren). Its aim is to initiate the larger project of substantiating our approach in full, which far exceeds the scope of this paper.

Upon acquaintance with the presented material, background sections could be skipped. Upon a thorough, orderly reading, boxed summaries could be skipped.
2  The Cogni-Gene’s Eye View

People have been struggling with love and trying to figure it out for quite some time now. This work’s immense pretension, in shedding fresh valuable light on it through a shifted perspective on agenthood, is constrained only by the fact that almost everything has already been done for us. The ground has been prepared by a much more general revolution in the understanding of evolution and life (its product) itself, with the philosophical grandeur emanating from scientific progress (with its healthy empirical grounding and mathematical grip). A few decades later, science is still only beginning to uncover the complex connections between actual genes and their physiological phenotypic products. And as for their behavioral products, the challenge is much harder, as these are mediated through brain mechanisms, an ultra-complex connection of its own that science has only begun to explore. But the most-general principles are better-understood, and do bring order and allow for some novel explanations and further predictions. This conceptual revolution, of the gene’s eye view, will serve us not only as a general inspiration but concretely for setting up the underlying genetic (evolutionary) level.

Let us note here from the outset that its task will be two-fold:

- Making sense of and making room for unifications of individuals – at this underlying level (of the genes that orchestrate our construction into what we are).
- Bringing in normativity, constituting the interests (the game-theoretic utilities or preferences) of the gene-agents and unions thereof – through evolution (natural selection determining and constituting what we are for).

2.1  Background: Dawkins

Dawkins’ early works did a fine job of summing up and getting to the conceptual heart of this revolution in evolutionary thinking, which has since taken over (much because of these works). Here we present their most abstract core.

2.1.1  The Gene’s Eye-View

That central conceptual revolution in evolutionary theory has been the transition\(^8\) to the gene’s eye view, as captured early on by Dawkins (1976) (following Williams (1966)). To recapitulate:

Consider a gene, as we shall always do here, as an 'atomic' region of code (namely within a complete genome), small enough for all intents and purposes\(^9\). And an allele of that gene as a particular code (though we shall generally adopt the common imprecise language of using ‘genes’ to

--

\(^8\) The picture is of course not as simple, and there are some prominent opponents of this view.

\(^9\) The precise concept of a gene continues to be refined. See discussions in Williams and in Dawkins’s, as well as [https://en.wikipedia.org/wiki/Gene#Modern_evolutionary_synthesis](https://en.wikipedia.org/wiki/Gene#Modern_evolutionary_synthesis) etc.
designate alleles, too). The set of alleles of a gene will thus be its available strategies. Importantly, alleles & genes must be considered not as specific tokens of entities (which are associated with some particular organism), but rather as abstract types of entities (which different organisms can share).

Viewed this way, a gene’s fate differs categorically from the fate of an organism which possesses it. That is, their utilities are different; their interests diverge on principle. This conceptual shift makes room for (organism-)altruism (through kin-selection) within a mechanistic framework where only ‘cold-blooded’ rationality plays a role (‘rationality’ provocatively amounting to selfishness). The focus is now on the gene that needs to be selfish if it is to be found in nature – not on the organism.

This property of ’mechanistic explanation’ is actually an acute conceptual matter. While Darwin himself was perfectly careful in his writings, a large school-of-thought seemingly within what developed to be ‘the theory of evolution’ was actually sticking to the old, upside-down picture of Aristotelian teleology. In this picture, the historical process ‘strives’ towards some goal; namely towards ‘survival of species’ (not to mention the bluntly ethnocentric ‘human race’). Evolution becomes the purpose rather than an emergent by-product of the process of natural, blind selection. This, specifically in the form of group-selection, is a major fallacy in evolutionary thinking10. A fallacy that concerns the very heart of what a scientific explanation is.

Once one learns to see the (old-style, naïve group-selection) fallacy for what it is, the gene’s eye view takes its place as the anti-group-selection reasoning brought to its logical extreme. Even single organisms are in fact, in the important sense, groups; groups of embodied genes. And genes, as pieces of code, are ultimately the true atoms the evolutionary process operates upon. A coded instruction, an action, that is better fitted with respect to its environment – its neighboring coded instructions included – will prosper, while less fitted ones will die out.

2.1.2 Organisms as Pseudo-Agents

We accept (for the rest of this work) the gene’s eye-view of evolution. But surely the original conception, looking upon the level of the individual organism as what natural selection acts upon, couldn’t have been completely misdirected? After all, its eminence resulted from its successfully putting much of the biodiversity into a single, coherent, structured picture – specific unresolved issues such as altruism notwithstanding.

The key point is this: a gene usually has no "independent, unmediated reach into reality". It is the organism that serves as a ‘vehicle’, allowing for the gene to survive and make new copies of itself. Now, as for the mutual genes that share this same

---

10 Besides the large efforts spent by Dawkins (1976), it receives a more explicit philosophical criticism by Dennett (1995). How mechanistic, valid forms of group-selection do possibly fit into our framework, will be discussed in Part II.
vehicle, their utilities coincide in most respects. But moreover, in all issues where other copies of the same genes are not concerned (that is, excluding sexual reproduction, kin selection and the likes), the utilities of the individual genes coincide with those of their vehicle, the organism; whatever is good for them is good for it, and vice versa. The identification of utilities entails that, as far as such evolutionary or game-theoretic analyses of interactions are concerned, there is no competition; the individual agents can conceptually be unified into a single, 'bigger' agent (whose possible actions comprise of the combinations of the smaller, atomic agents); a conglomerate of different units acting in harmony. This is what’s behind the ability of the organism-level evolutionary perspective to explain so much.

The general philosophical issue for us is that the latter evolutionary identification between an organism and its genes is a delicate, contingent matter, not at all an a priori truism. Breaking the agenthood, the utility, down to the new atoms (each gene’s own utility), and reassembling it, is just one aspect of this. A further central difference between an organism and its genes’ utilities, which we’ve put aside (until chapter 6), is of course where other copies of the genes are concerned; as when an animal invests in its offspring at its own expense. One remaining central aspect concerns the very status of the organism as the product of its genome:

At the most abstract, information-theoretic core of the evolutionary process, the choice of looking upon the organism as the genes’ vehicle is a somewhat-arbitrary one, motivated by "merely-biological" facts. The organism – as a (rather than the) level of the phenotypic expression of the genotype, as a compound material object that is strictly separated from its environment, thereby making up for a certain division of the world – is simply not the only "true" choice. It certainly plays a central role in many types of evolutionary analyses, but there are others.

The central work dedicated to making this point and giving it its due biological substance is Dawkins’ (1989). It will be necessary for us in defusing The Central Dogma, as we are to consider an entity ranging over two organisms.

For a fairly simple example to start with, consider a bird’s nest (a beaver’s dam is his favorite more provocative example). Although not ‘biologically’ a part of the parents’ body, it serves the purpose of providing for a safe place for the little ones to stay at until they reach sufficient mobility and independence just as much as the kangaroo’s pouch does. And the propensity to build such a nest is just as much of a genetically-programmed evolutionary adaptation. The limits of the body are not the essence.

---

11 One can conceive of an allele that can mildly hurt the organism while helping him recognize and favor fellow organisms that carry that same allele, thereby spreading within the population at the expense of his fellow passengers. But such self-recognition is quite sophisticated and is not what almost all genes come close to controlling.
For a deeper example, consider a parasitic Gordian worm that, rather than simply make use of its cricket host as a supply of resources, at the end manipulates the cricket into killing itself by jumping into the water – from where the worm breaks out to carry on with its life-cycle. That worm is genetically programmed to manipulate its environment for its own sake – other organisms’ brains included. As Dawkins summarizes –

The Central Theorem of the Extended Phenotype:

An animal’s behavior tends to maximize the survival of the genes "for" that behavior, whether or not those genes happen to be in the body of the particular animal performing it.

The only thing genes influence directly is the synthesis of proteins. But as for indirect influences, they can, and do, go much beyond the organism’s own body, principally through behavior as shaped by cognitive machinery. And some phenomena simply require this broadening of context, going beyond the organism as “the” phenotypic expression of a genome, in order to be understood in the right evolutionary context. Love, as we shall posit, is one such phenomenon – involving coalitional interaction between genes of different organisms.

The fact that we understand so very little of brains’ inner-workings and the way they are ultimately brought about by the genes (far less than the standard in biology) does not completely rule out the possibility of successfully reaching any sort of related understanding. And cognition as at the service of the genes shall indeed be what we will need to actually say something novel about in this work. But before we move on to the finer complexities of cognition, let us complete the break away from the organism-centered evolutionary view.

2.1.3 Symbiosis

A wide range of interactions between organisms is analyzed in "The Extended Phenotype". We generalize over them, giving a central, constitutive role to (contextualized) fate-sharing as determining cooperation.

Although the two former examples adhere to the extending of the phenotypic reach of the genome beyond the organism, that genome’s genes themselves are still grouped together according to the single organism to which they belong. E.g., the parasitic worm itself can still be considered to be taking after its, its genome’s, interests – as opposed to the cricket’s. The organism as its genes’ vehicle, still enforces a common fate for them, equating their utilities, and opposing these with the utilities of the genes of their rivaling organism.

But let us now consider the opposite, (known as "obligate mutualistic") symbiotic state, where two organisms completely depend on each other for survival.

Take for example fig wasps, which rely exclusively on figs to supply them not only with food and shelter, but with their exclusive mating spot, and in return exclusively pollinate the figs. Although biologically perfectly-distinct species, as their lives are tied up together, so are their evolutionary courses. Indeed, very specific types of fig wasps specialize in very specific types of
FIGS, THE TWO CO-EVOLVING TOGETHER AS SPECIES. THE MANY INTERDEPENDENCIES PUT SERIOUS CONSTRAINTS ON THE POSSIBLE EVOLUTIONARY COURSES OF EACH OF THE SPECIES AS A SEPARATE ONE, TO THE POINT WHERE THE SEPARATION INTO TWO DISTINCT BIOLOGICAL ENTITIES LOSES A BIT OF ITS EVOLUTIONARY HOLD.

YET WE MAY STILL TEND TO THINK OF THIS AS CO-EVOLUTION, AS INTERACTION BETWEEN THE TWO SPECIES AS THE AGENTS, RESPONDING TO ONE ANOTHER’S CHANGES ON AN EVOLUTIONARY SCALE (THE SAME WAY A PREDATOR AND ITS PREY MAY BOTH BE FORCED TO SPEED UP), RATHER THAN AS EVOLUTION OF A SINGLE SPECIES. THE REASON IS THAT THE LIFECYCLES OF THE TWO ARE NOT PERFECTLY INTERTWINED, AND SO SO ARE THEIR FATES: A MUTATION IN A WASP’S GENE THAT WOULD BENEFIT THE FIG, WOULD MOST LIKELY NOT BENEFIT THAT VERY SAME MUTATED WASP IN RETURN; SAY, IF IT GIVES THE OTHER WASPS THAT SHARE THAT FIG AN ADVANTAGE OVER IT; OR IF IT HELPS THE FIG ONLY IN THE LONG RUN, LONG AFTER THE WASP HAS DIED OR LEFT THAT FIG. AND SO NATURAL SELECTION STILL DIFFERENTIATES BETWEEN THE GENES OF THE TWO GENOMES; THEIR PRIVATE UTILITIES STILL DIFFER A LOT.

Perhaps the most famous biological example of commonality-of-interest so extreme as to challenge the distinction of the organism as the individual, are the social insects (Dawkins, 1976 pp. 171-). The genetic infrastructure makes it, for the most part, more worthwhile to "forget about yourself", leave the task of reproduction to the queen, and just take after the colony, alongside everyone else in it. The comradeship is strong enough so as to be expressed not only behaviorally, but physiologically, as with an in-built weapon whose usage means personal death. So little need for autonomous self-sufficiency motivates specializations, to the point that an individual or group of individuals take roles akin more to body-parts than to individuated organisms. A whole colony is thus best thought of as a super-organism. This applies just as much to its cognitive functioning. At the lower, organism level, a bee would be willing to commit suicide (and we can only begin to speculate "What Is it Like to Be a Bee"). Or share information about the direction to pollen etc.12, but give up the persuasion cooperatively, if another brings alternative information, more valuable for the whole. Through the interaction of the many, emerges a far richer cognitive system, serving the colony, creating something that may be argued to come closer to our notion of an 'individual' (which in particular wouldn't be expected to naturally commit suicide).13 Of course, various genetically-driven conflicts do arise even in kinship (just like parent-offspring conflicts), and so do create a less-than-perfect picture of harmony. But the differences from non-social animals illuminate the general principles.

Mutualistic and even obligate symbiosis comes in many shades, then. It may be dimmed not only by one side – as a species – having a choice to cooperate with one of a few (therefore competing) species, but by one side – as an organism – not completely depending on a single matching organism as cooperator. A further important loosening is the possibility of context-dependent cooperation. Time-dependent, for example: At night, one side may provide required heat and the other protection, while having absolutely nothing to do with each other during the day.

12 https://en.wikipedia.org/wiki/Waggle_dance
13 This topic is nowadays a rich & active area of research, engineeringly (in robotics) and scientifically (e.g. http://www.weizmann.ac.il/complex/feinerman/).
2.2 Extending the Extended Phenotype

What comes out of the many examples of (different levels of) symbiosis can be put more generally, at the level of the genes, as –

**The Cooperation Maxim:**

*The potential for symbiosis is directly proportional to the degree of common fate sharing.*

Bound within the same organism, one gene’s interest to help a fellow one (in almost all cases\(^\text{14}\)) is just as great as the interest to help itself (as these are one and the same), and so we get perfect harmony. But genes, even if belonging to different genomes, may just as well find themselves in a situation where the interest to cooperate is just as strong.

Perhaps context-dependent, perhaps temporary, but still, given that situation, just as strong. The more all-embracing that context is, the more meaningful it becomes to consider the shared phenotypic entity as a single, multigenomic organism in general. But the pervasiveness of the context aside, within that particular context, it is *always* a priori meaningful to view all genes whose utilities within that context coincide as a single entity; a grouping that makes sense *relative to that specific context.*

This principle might to some extent be concealed in classical, biological evolutionary theory (as opposed to evolutionary psychology). This is due to the focus on the organism (modulo its cognition) level of the phenotypic expression – a level at which the genes’ product is a much more ‘hardwired’, context-free creation, relatively-fixed throughout the many different states along the life of the organism and its genes. The restriction to this specific level simplifies and therefore enables the exploration of many issues; But many others are simply too delicate for this approximation, *having to do with genes’ adaptation to the actual dynamics of life, rather than simply being the best single response in expectancy over all states in life.* The common biological view is led to this approximation by the conception of the gene’s phenotypic expression as a building instruction, which is thus seemingly a context-free result.\(^\text{15}\) But genes *can* in principle influence in a context-dependent fashion.

This is where *cognition* (in the most general sense of the word) comes in.

---

\(^{14}\) Bar Selfish Genes.

\(^{15}\) All this is a highly simplistic portrait of biology, to be sure, here only for the purpose of stressing the issue of context-dependence. Outside of cognition, there is still much in the biological world that is context-dependent, involving environmental inputs & outputs mediated by some nontrivial mechanism, even if this happens to be a physiological, local mechanism rather than neurological brain operated. e.g. trees shedding leaves and so many other seasonal or other time-dependent phenomena.
2.2.1 Cogni-Genes

Consider, to begin with, an organism (as a manifestation of its genes’ consortium) in a primordial state of cognitive development, capable of taking care of itself evolutionarily, by surviving and reproducing (no need for much – trees can spread their seeds rather randomly with the wind, for example). There is the obvious pressure to develop more sophisticated cognitive machinery (though, also obviously, constrained by its cost) – sheerly for the sake of that organism. A better cognitive representation of the world (brought about by some genes) allows for more exact, context-dependent behavior (quite possibly brought about by other genes).

One gene may endow the organism with some no-cost, very simple cognition capability, or it may even be a trivial one (say, local, crude temperature "recognition", simply by cells being influenced by it). Then, given a pre-existing such basic cognition capacity, a different gene may be able to make use of it (for its own sake, regardless of the other genes’ matching interests), by adding some useful, context-dependent behavior. Thereafter, given this already-fixed basic context-dependent behavior that is in favor of the organism’s sake, there may be an advantage in simply improving those underlying cognition abilities, as this could make the exertion of that behavior more precise. And so on.

Dealing with the immensely complicated issue of the evolution of cognition is not at all our intention here. For our purpose, all we need is the acceptance that the general principles of evolution apply to it just as fiercely. Genes are selected for their value as locally-advantageous patches. Value — for themselves. There is no predetermined goal in terms of survival of a group. No group. Not of ‘individual’ organisms, nor of those genes belonging to a single organism. The apparent harmonious functioning of the collective is not an a priori necessity, but a consequence of being in the same boat; an emergent property. Just as this applies to how genes bring about the physical body of the organism (modulo its brain), it applies to the brain itself, and so to its product — the organism’s mind.

The mind is thus, to begin with conceptually, a multitude of great many different cognitive (multi-leveled) facilities or patches; ultimately, the phenotypic expressions of their matching genes. Each such ‘cogni-gene’ is selected for doing best what it does, independently of all others (not just those belonging to different genomes). The others’ manifestation constitutes the full cognitive framework within which it operates; its immediate environment.

We’ve already learned to overcome the dogmas of the organism as the only meaningful, the "correct", level of phenotypic expression, and its genome as the only meaningful, the "true", grouping of genes. We now need apply that lesson in particular to those ‘cogni-genes’, with the analog to the full organism being the particular phenotypic expression of the cogni-genes – the

---

16 The fact that most of earth’s biomass exhibits very little cognitive sophistication (if at all) may be the result of a combination of various factors: Size constraints (a general tendency towards being much smaller than humans, which also constrains brain-size), other biological constraints (e.g. computational architectures like neuron-based ones unavailable to plants), as well as evolutionary ones (accumulation of advantageous mutations may be much harder in regard to cognitive systems. Algorithm-patches do not add up as nicely as physical constructs).
self (a.k.a. the 'I', the 'soul'). Renouncing its immanent reality, its distinct metaphysical status within the realm of the mental, its defining stature as the object of exploration of psychology at large, is not simply counter-intuitive at first. It clashes with the way we are brought up to think of our-selves (at least in Western culture); it may go against our very being (so to speak). This level will be the subject of the next chapter. But for now let us focus on the genes’ direct view of things, as a break away from the organism’s.

As we’ve discussed, then, for a cognitively-simple organism, there’s pressure to further develop its cognitive capabilities — for its own sake. Initially, the genes’ interests are factored, channeled, through the genome in which they partake: As a cogni-gene can hardly tell apart and react differently to the different contexts that occur within its organism’s lifetime, the only coalition it can form, the only true alliance it can count on — on average — is with its fellow genes, who share the same organism, and conjointly (aim to) transfer on to new ones.

To try and illuminate this point through an exception: A womb–sharing embryo should be built as to not make the full, greedy use of available resources with disregard to how little extra value it has to gain. Not because of its own sake, and beyond the sake of its bearer — for the sake of its womb-mates’, who partly share its genes. As the body is built gradually, genes indeed get this indirect, rather simple sort of cognition capability. They recognize a stage in the developmental timeline (which is an important but very partial component of the full context-recognition throughout life), a stage when they should be in ‘kin-selection mode’. And so appropriately act more altruistically (and can accept the same from their peers). But barring such exceptions, each organism’s genes can ordinarily count on cooperation mostly amongst each other.

In order for the organism to serve its genes well, beyond simply taking after its own best interest as an entity in itself (i.e. survival), they need to build a sex drive into it; enforcing other biases (relative to its own interests) such as a need to feed its young may also help. But the organism’s cognitive system may still function as a whole, helping it pursue its many dictated goals — and nothing else. An elderly saving up more money than he’d ever have the chance to spend needn’t be deemed irrational, but rather simply rational relative to his bias towards helping his descendants even after he’s gone.

But now: The richer the cognitive representation of reality, the finer and more meaningful its partition into different contexts, the better the cogni-genes may be able to act, by taking the best action in each context — rather than just the best action in general. In particular, with information-processing capabilities sufficient for individuating particular other genes or genomes (e.g. through recognizing a spouse), the best action may involve coordination with these’ own actions. And ‘the best action’, the one selected-for by evolution, is ultimately that which is best for the cogni-gene’s own benefit — not for its organism’s.

What follows from this is the conclusion we required of the genetic level -

**Trans-Genomic Cooperation:**

A cogni-gene acting within a cognitive framework sufficient for setting apart situations in which its interests coincide with those of genes beyond its own genome, may be able to partake in an extended such coalition.

---

17 as always, through its cognitive manifestation
Dawkins' (1989) logic was to break down the organism into a multitude of gene agents, and 'rediscover' it as an entity. An entity that is no longer atomic, but instead a coalitional unity that opted to be forced to stick together, bringing it into harmony. A harmony that is a product of their situation rather than the entity's a priori (metaphysical) essence. We offer Trans-Genomic Cooperation as the extension of this logic: Along with the rest of the body as a product of the genes, their mindly product (through their bodily product) of the self, too, gets to be 'rediscovered' as a harmonic unity that is the result of fate-sharing rather than a priori essence. But unlike the physical body, it can in principle be flexible enough to respond to converging interests of genes beyond the single genome. A bigger resulting 'self' would thus be not inherently different form the classic, individualistic self – just the unity of a bigger coalition.

That this may all actually manifest itself at the psychological level and how, and when – will be our concern in what follows.
Chapter Summary

The psychological level: Enter *The Self*. First as a general topic, and then within a particular framework, as a "head of state". This framework makes room for unions into bigger selves that are not inherently different.

The revolution that was the gene’s eye view, now applied to cogni-genes as in the previous chapter, suggests that we may have to revolutionize our conception of their own particular mindly product, the self. That is, if we would have a fairly clear agreed-upon conception to begin with – which is not quite the case. Merely making precise our pre-theoretic notion isn’t straightforward. And beyond naïve conceptions, on the theoretical front, the nature of the self is a central, hotly-debated (and rich and complicated and highly speculative) topic in the philosophy of mind.

Following an introduction to the topic, we bring in an existing framework introduced by Humphrey & Dennett (1989) (or rather, let the authors do). This work is in the time-honored tradition of studying the way our brains and minds work, through close inspection of *how they fail to* – when they do. Thus, the motivations for this framework are independent of ours – but it will serve us well for our psychological level. It is (as always) worth keeping in mind what we are after: the possibility of uniting to form a bigger self (one that is not of an inherently different nature).

3.1 Background: Prelude to Philosophy of the Self

Humphrey & Dennett’s own introduction to the topic, in preparation for their framework.

Many people who find it convenient or compelling to talk about the "self" would prefer not to be asked the emperor’s-new-clothes question: just what, exactly, is a "self"? When confronted by an issue that seems embarrassingly metaphysical, it is tempting to temporise and wave one’s hands: "It’s not a thing, exactly, but more a sort of, well, a concept or an organising principle or . . ." This will not do. And yet what will?

Two extreme views can be and have been taken. Ask a layman what he thinks a self is, and his unreflecting answer will probably be that a person’s self is indeed some kind of real thing:

---

18 As the framework is a very particular & peculiar suggestion rather than fairly common background which the reader may be expected to perhaps already be familiar with, we will bring the original authors’ relevant text in the full here.

19 The recommended alternative is to simply read the paper itself. Also by the authors and highly relevant are (Dennett, 1991; Dennett, 1992; Humphrey, 2000).
a ghostly supervisor who lives inside his head, the thinker of his thoughts, the repository of his memories, the holder of his values, his conscious inner "I". Although he might be unlikely these days to use the term "soul", it would be very much the age-old conception of the soul that he would have in mind. A self (or soul) is an existent entity with executive powers over the body and its own enduring qualities. Let's call this realist picture of the self, the idea of a "proper-self".

Contrast it however with the revisionist picture of the self which has become popular among certain psychoanalysts and philosophers of mind. On this view, selves are not things at all, but instead are explanatory fictions. Nobody really has a soul-like agency inside them: we just find it useful to imagine the existence of this conscious inner "I" when we try to account for their behaviour (and, in our own case, our private stream of consciousness). We might say indeed that the self is rather like the "centre of narrative gravity" of a set of biographical events and tendencies; but, as with a centre of physical gravity, there's really no such thing (with mass or shape or colour). Let's call this non-realist picture of the self, the idea of a "fictive-self".

Now maybe (one might think) it is just a matter of the level of description: the plain man's proper-self corresponds to the intrinsic reality, while the philosopher's fictive-selves correspond to people's (necessarily inadequate) attempts to grasp that intrinsic reality. So, for example, there is indeed a proper-Nicholas-Humphrey-self that actually resides inside one of the authors of this essay, and alongside it there are the various fictive-Humphrey-selves that he and his acquaintances have reconstructed: Humphrey as seen by Humphrey, Humphrey as seen by Dennett, Humphrey as seen by Humphrey's mother, and so on.

This suggestion, however, would miss the point of the revisionist critique. The revisionist case is that, to repeat, there really is no proper-self: none of the fictive-Humphrey-selves – including Humphrey's own first-hand version – corresponds to anything that actually exists in Humphrey's head.

At first sight this may not seem reasonable. Granted that whatever is inside the head might be difficult to observe, and granted also that it might be a mistake to talk about a "ghostly supervisor", none the less there surely has to be some kind of a supervisor in there: a supervisory brain program, a central controller, or whatever. How else could anybody function – as most people clearly do function – as a purposeful and relatively well-integrated agent?

*The answer that is emerging from both biology and Artificial Intelligence is that complex systems can in fact function in what seems to be a thoroughly "purposeful and integrated" way simply by having lots of subsystems doing their own thing without any central supervision.* Indeed most systems on earth that appear to have central controllers (and are usefully described as having them) do not. The behaviour of a termite colony provides a wonderful example of it. The colony as a whole builds elaborate mounds, gets to know its territory, organises foraging expeditions, sends out raiding parties against other colonies, and so on. The group cohesion and coordination is so remarkable that hard-headed observers have been led to postulate the existence of a colony's "group soul" (vide Marais' "The Soul of the White Ant"). Yet in fact all this group wisdom results from nothing other than myriads of individual termites, specialised as several different castes, going about their individual business – influenced by each other, but quite uninfluenced by any master-plan.
3.2 Background: Identity and Head of State

Section Summary

Humphrey & Dennett introduce their framework, through which they attempt to explain "Multiple Personality Disorder".

The self is likened to the head of state (and, in particular, personality formation to elections).

What the authors set out to explain is Multiple Personality Disorder, known more academically as "Dissociative Identity Disorder" (it is notoriously under dispute, but we need it mainly for illustrative purposes, so bear with us). The disorder may be familiar to the reader (hopefully only) from the many films that exploit its dramatic potential. It involves a single body hosting in practice different persons at different times. Not only does the personality differ dramatically (possibly including handwriting, dressing-style, etc.), but the identity of the self itself: Who she thinks she is and what she remembers (general knowledge is shared between the personalities, particular memories are not. Hence, she might be able to play guitar without knowing how she does). And as the mental state is not at all disconnected from the physical one, beyond EEG-detectable differences, the different egos may actually exhibit different physiological ones (regarding sensitivity to drugs, for example).

The disorder seems to be highly related to early-childhood sexual abuse. Portraying a typical such case, the authors raise the possibility of the multiplicity as a "division of emotional labour ... not only.. viable, but.. rewarding all around." They then go on to present a very general (if sketchy and highly metaphoric) framework for the subject of selfhood, within which the disorder fits naturally. As this framework is central to our own work, and as it may perhaps be quite uneasy to digest at first, let us continue with the excerpt of their own description (Humphrey, et al., 1989) in full (again, with our emphasis):

Then is the argument between the realists and the revisionists being won hands down by the revisionists? No, not completely. Something (some thing?) is missing here. But the question of what the "missing something" is, is being hotly debated by cognitive scientists in terms that have become increasingly abstruse. Fortunately we can avoid – maybe even leapfrog – much of the technical discussion by the use of an illustrative metaphor (reminiscent of Plato's Republic, but put to quite a different use).

Consider the United States of America. At the fictive level there is surely nothing wrong with personifying the USA and talking about it (rather like the termite colony) as if it had an inner self. The USA has memories, feelings, likes and dislikes, hopes, talents, and so on. It hates Communism, is haunted by the memory of Vietnam, is scientifically creative, socially clumsy, somewhat given to self-righteousness, rather sentimental. But does that mean [here is the revisionist speaking] there is one central agency inside the USA which embodies all those qualities? Of course not. There is, as it happens, a specific area of the country where much of it comes together. But go to Washington and ask to speak to Mr American Self, and you'd find there was nobody home: instead you'd find a lot of different agencies (the Defense Department, the Treasury, the courts, the Library of Congress, the National Science Foundation, etc.) operating in relative independence of each other.

To be sure [and now it is the realist speaking], there is no such thing as Mr American Self, but as a matter of fact there is in every country on earth a Head of State: a President, Queen, Chancellor, or some such figurehead. The Head of State may actually be non-executive;

---

20 Which is "congenial to Dennett’s] way of thinking about the self, and the social milieu in which a self is required." (Personal communication)
certainly he does not himself enact all the subsidiary roles (the US President does not bear arms, sit in the courts, play baseball, or travel to the Moon ..). But nevertheless he is expected at the very least to take an active interest in all these national pursuits. The President is meant to appreciate better than anyone the "State of the Union". He is meant to represent different parts of the nation to each other, and to inculcate a common value system. Moreover – and this is most important – he is the "spokesman" when it comes to dealing with other nation states.

That is not to say that a nation, lacking such a figurehead, would cease to function day-to-day. But it is to say that in the longer term it may function much better if it does have one. Indeed a good case can be made that nations, unlike termite colonies, require this kind of figurehead as a condition of their political survival – especially given the complexity of international affairs.

The drift of this analogy is obvious. In short, a human being too may need an inner figurehead – especially given the complexities of human social life. Consider, for example, the living body known as Daniel Dennett. If we were to look around inside his brain for a Chief Executive Module, with all the various mental properties we attribute to Dennett himself, we would be disappointed. None the less, were we to interact with Dennett on a social plane, both we and he would soon find it essential to recognise someone – some figurehead – as his spokesman and indeed his leader. Thus we come back full circle, though a little lower down, to the idea of a proper-self: not a ghostly supervisor, but something more like a "Head of Mind" with a real, if limited, causal role to play in representing the person to himself and to the world.

The authors aren't explicit about this, but we suggest thinking of this head of state (who's in the spotlight) as the conscious self, with everything below him in the governmental hierarchy and the many citizens as the unconscious part. And so the state as a whole constitutes the full mental being. This seems to be in line with their text and intentions, and will fit our own story. If this is accepted (as we think it should be), we can turn to the vexed question of self-development or self-establishment. Here the Head of State analogy may seem at first less helpful. For one thing, in the USA at least, the President is democratically elected by the population. For another, the candidates for the presidency are pre-formed entities, already waiting in the wings.

Yet is this really so? It could equally be argued that the presidential candidates, rather than being pre-formed, are actually brought into being – through a narrative dialectical process – by the very population to which they offer their services as president. Thus the population (or the news media) first try out various fictive versions of what they think their "ideal president" should be, and then the candidates adapt themselves as best they can to fill the bill. To the extent that there is more than one dominant fiction about "what it means to be American", different candidates mould themselves in different ways. But in the end only one can be elected – and he will of course claim to speak for the whole nation.

In very much a parallel way, we suggest, a human being first creates – unconsciously – one or more ideal fictive-selves and then elects the best supported of these into office as her Head of Mind. A significant difference in the human case, however, is that there is likely to be considerably more outside influence. Parents, friends, and even enemies may all contribute to the image of "what it means to be me", as well as – and maybe over and above – the internal news media. Daddy, for example, might lean on the growing child to impose an invasive fictive-self.
Thus a human being does not start out as single or as multiple – she starts out without any Head of Mind at all. In the normal course of development, she slowly gets acquainted with the various possibilities of selfhood that "make sense" – partly through her own observation, partly through outside influence. In most cases a majority view emerges, strongly favouring one version of "the real me", and it is that version which in installed as her elected Head of Mind. But in some cases the competing fictive-selves are so equally balanced, or different constituencies within her are so unwilling to accept the result of the election, that constitutional chaos reigns – and there are snap elections (or coups d'état) all the time.

… We have talked here in metaphors. But translations into the terms of current cognitive science would not be difficult to formulate...

In a later paper (2000), Humphrey takes on this much more fundamental concern of personality and consciousness formation through childhood, connecting it as we also will to the issue of unity\textsuperscript{21} (emphasis is his):

\begin{quote}
[P]arts come to belong to a whole just in so far as they are participants in a common project..
\end{quote}

What makes the parts of a person belong together – if and when they do? The clear answer has to be that the parts will and do belong together just in so far as they are involved in the common project of creating that person’s life…

I may indeed be made up of many separate sub-selves, but these selves have come to belong together as the one Self that I am because they are engaged in one and the same enterprise: the enterprise of steering me – body and soul – through the physical and social world. Within this larger enterprise each of my selves may indeed be doing its own thing: providing me with sensory information, with intelligence, with past knowledge, goals, judgements, initiatives, and so on. But the point – the wonderful point – is that each self doing its own thing shares a final common path with all the other selves doing their own things. And it is for this reason that these selves are all mine, and for this reason that their experiences are all my experiences. In short, my selves have become co-conscious through collaboration…

the baby boy is beginning to come together. Already there are hints of small collaborative projects getting under way: his eyes and his hands working together, his face and his voice, his mouth and his tummy. As time goes by, some of these mini-projects will succeed; others will be abandoned. But inexorably over days and weeks and months he will become one coordinated centrally conscious human being. I begin to understand how in fact he may be going to achieve this miracle of unification. It will not be.. through the power of a supervisory Self who emerges from nowhere and takes control, but through the power inherent in all his sub-selves for, literally, their own self-organisation.

Everything quoted here will apply perfectly to our case, which would involve an enterprise much truer than just "steering me – body and soul – through the physical and social world" (which is but a proxy). And certainly much more central than Multiple Personality Disorder. Their framework, granted, is all highly-speculative, even as the rough preliminary metaphor that it is meant to be. Their motivation for this philosophical framework is the way in which it fits, helps explain the disorder, and make it seem more ‘natural’, thereby reinforcing the reasonableness of the framework itself. The reader should not expect to be reasonably-convinced without at

\textsuperscript{21} His independent (but earlier) development of this theme and connection to personhood consists in but a few sketchy remarks, and confuses elements which our own more extensive development in the next chapters sorts out (namely, coordination-problems vs. utility-collisions).
least having read the rest of their discussion. This issue, of the relation between the framework and the disorder, is not directly of our concern here. But extending its logic – is. And part of their attempt to sugar the pill and justify their suggestion, beyond the fact that we are relying here on the reasonableness of the framework, will lend itself to this.

3.3 Personal Identity Generalized

The authors' persuasion for their central suggestion is used, spelling out how its logic extends naturally to give us what we need for the psychological level.

Multiple Personality Disorder aside, the authors’ framework will prove to be instrumental to our own agenda, which, recall, in a sense concerns the opposite direction: multiple (well, two) bodies sharing a single self instead of a single body sharing multiple selves. The principles are similar. Let us then end the long excerpt with the part of their justification (1989) that relates directly to our business (as usual, with our emphasis), drawing out the generalizable:

There can… be no guarantee that either the speaker or anyone else who hears him over an extended period will settle on there being just a single “I”. Suppose, at different times, different subsystems within the brain produce “clusters” of speech that simply cannot easily be interpreted as the output of a single self. Then – as a Bible scholar may discover when working on the authorship of what is putatively a single-authored text – it may turn out that the clusters make best sense when attributed to different selves.

This logic works both ways: just like it may turn out that clusters of speech (and more generally, actions) make best sense when attributed to different selves, despite originating from the same mouth (or more generally, body), actions may make best sense when attributed to a single self – despite originating from different bodies. If we are willing to let go of a priori per-body attribution of agenthood (as the revolution of the gene’s eye view would have us do), if agenthood is a more dynamic and complex phenomenon that requires attention, then this logic is inescapable. We will breathe more life into this in the next chapter. Carrying on with their justification:

How about the selective amnesia shown by different Heads of Mind? To readers who have even a passing knowledge of computer information processing, the idea of mutually inaccessible “directories” of stored information will already be familiar. In cognitive psychology, new discoveries about state-dependent learning and other evidence of modularisation in the brain, have led people to recognise that failure of access between different subsystems is the norm rather than the exception. Indeed the old Cartesian picture of the mind “transparent to itself” now appears to be rarely if ever achievable (or even desirable) in practice. In this context the out-of-touchness of different selves no longer looks so startling…

But what should in this context look even more unstartling is the literal out-of-touchness of different brains – under the possibility of sharing part in a larger, unified computation and the agenthood it brings about. As even different modules regularly have no access to each other and communicate only through very simple, limited interfaces, mere physical disconnectedness

---

22 Further, modern or more popular justifications, could run along the lines of:  
Dennett’s "Multiple Drafts Model" ([http://plato.stanford.edu/entries/consciousness/#NarIntThe](http://plato.stanford.edu/entries/consciousness/#NarIntThe))  
Information Integration Theory: [http://plato.stanford.edu/entries/consciousness/#InfIntThe](http://plato.stanford.edu/entries/consciousness/#InfIntThe)
between different brains alone does not automatically entail a mental separation. First off, the separate brains can still communicate, even if quite limitedly. Information is an abstract entity, that transcends rigid physical boundaries. And furthermore, the decentralization principle mentioned, where "thoroughly 'purposeful and integrated' functioning can result from subsystems doing their own thing without any central supervision", implies that none of the brains has to in any way be privileged. None has to accommodate that Cartesian Theater\textsuperscript{23} where everything appears before that self’s consciousness. (We will revisit these themes once our full framework is in place and love’s place in it has been explicated).

Last but not least, is their more dynamic view of personhood, which can in certain cases respond to life’s happenings quite drastically and abruptly:

Even the idea of an "election" of the current Head of Mind is not implausible. Events very like elections take place in the brain all the time – whenever coherent patterns of activity compete for control of the same network. Consider what happens, for example, when the visual system receives two conflicting images at the two eyes. First there is an attempt at fusion; but if this proves to be unstable, "binocular rivalry" results, with the input from one eye completely taking over while the other is suppressed. Thus we already have, at the level of visual neurophysiology, clear evidence of the mind’s general preference for single-mindedness over completeness.

As they say, ”in the human case.. there is likely to be considerably more outside influence. Parents, friends, and even enemies may all contribute to the image of 'what it means to be me'” (which we will be pushing further). Their own case of selfhood shakeup, purportedly due to extreme emotional stress and there for “the division of emotional labour”, as a pathology, necessarily infringes into the speculative, much less scientific, and non-formalizable realms of psychiatry. Luckily for us, our own case, as it will turn out, is quite different. But first, we will need to concern ourself with the general principles underlying such dynamics, and fit this framework of theirs as our psychological level – upon the genetic one (of the previous chapter).
So to connect what we’ve seen to what we soon will –

The "Head of State" framework makes room for:

- a dynamical view of the self,
- not rigidly tied one-to-one to an organism,
- including a natural, efficient formation of a larger (not inherently different) "self",
- according to underlying interests,
- and in general enables a more 'game-theoretic' approach.

\textsuperscript{23} (Dennett, 1991)
4 Unification and its Dynamics

Chapter Summary
We present an abstract approach to agent unification (and disassembly) and the strategic principles governing it, and examine a few simpler cases (than love) of such 'harmony'.

Customarily when setting out to model and analyze a scenario of interaction, the agents themselves, with their utilities and available strategies, are the pre-determined, fixed, starting-point. But as we are to understand love as agent-unification (at the genetic substrate as well as at its psychological product), we cannot make do with a fixed framework. We must set out to account for unification itself as a game-theoretic issue, make sense of it in terms of utilities and strategies. Abstracting these fundamentals would then let us account for love as a strategic choice.

The single agent re-conceived as a possibly compound unity of many "smaller" or simpler but perfectly valid agents, brings with it the complementary notion of many agents acting as one, which we shall term harmony. As our basic, static approach here is crafted in the footsteps of chapter 2, the archetype is the genome, with its harmonious product, the organism. The more weakly related social insects and the rather harmonious super-organism they constitute are another example. A few more cases of harmony will later on illustrate further important aspects, on our way to understanding love, too, in such terms. But we also need to attend to the dynamics behind unification. Here, we will instead propose the basic considerations on our own (deferring making the connections to biological cases and theory to part II). Let us now sketch these figuratively, before moving on to the somewhat more formal setting.

To borrow and carry on Humphrey & Dennett’s (1989) central metaphor, then, including the lucky (if incidental) choice of the United States of America: Countries split and merge throughout history (at times bypassing geographic dictates). Such a change may be a complicated and pricy process, but the end result may be worth it, in terms of a new grouping that better unites citizens of similar interests. By splitting a previous union of subjects that were too dissimilar for the union to function effectively for the general benefit of all of them; or by bringing together subjects that could unite effectively, thereby gaining the extra power, efficiency, and other advantages of a larger entity:

Once each uniting side can trust the other to act within the best interest of the unity rather than its selfish self, guards can be put down and future cooperations assumed, and new possibilities open up. Multiplicity of mandatory facilities can be avoided, for example (if one side can operate a power plant more efficiently due to access to the sea, the other side can avoid setting up its own, without putting itself up for extortion later). On the other hand, anything that the sides could have done before the merging which was in their best interest as a unity, can still be done just as well. This phenomenon, where "the whole is greater than the sum of its parts" – is synergy.

Thus, although there surely is a cost to unity, there’s also a synergetic value. The question central to this tradeoff between liberty and unity is: By how much does the unity’s interest deviate from those of the constituents? For parties whose goals are similar enough, the synergetic value in unity may outweigh the price of pursuing the unified entity’s goals instead of their own.

Where parties stand relative to this (split/merge) cut-off point may shift dynamically. Changes in the extent of similarity could even occur on a temporary, context-dependent basis: In a
world-war, two allies whose fates are now (if only until war’s end) tied up, may enhance the sharing of information between their intelligence agencies; they may help each other even on non-military issues; their leaders may cooperate, meet and try to reach joint decisions; etc.

Furthermore, splitting or merging would ordinarily carry a considerable price. Thus, instead of a single cut-off point, there’s actually a region of status-quo stability, above which a unification (of currently-separated) should occur, and under which a split (of currently-unified) should.

This principled outlook on countries and international relations is of course extremely simplistic; The actual historical dynamics of nations certainly does not abide by any such hyper-democratic citizens-serving ideal. But the ideal is important as such, we maintain, and does capture some essence (rather than just a fable). The analogy serves to enliven and highlight the aspects we need to capture, about which let us now be more concrete. As always – within the general framework of non-cooperative game theory (the games being multi-stage). The prospects for a full formalization, as well making the connection to the existing related literature\(^{24}\), shall be attended to in part II.

4.1 A Union

A Union: a group of agents for which their contextualized utilities identify. That is:

- Their utilities are one and the same\(^ {25}\) – in the non-indexical sense: possibility P\(_1\) preferred over P\(_2\) by both N\(_1\) and N\(_2\) must be the very same, e.g. both preferring that N\(_2\) would find a banana, as opposed to each one preferring to find a banana himself.
- But are one and the same – relative to the subgame they are in.\(^ {26}\)

A Union would ideally be thought of as a single agent, of that same shared utility function, and whose set of actions is the product of the action sets of the agents that make it up (i.e. no actual change, just a unified point-of-view). Ideally – though some important caveats will soon follow. This approach (as a stance in the philosophy of game theory) thus posits Dawkins’ fate-sharing theme as a constitutive principle for when a multitude of agents could legitimately be considered as one.

This simplistic form will for the most part suffice for our illustrative purposes. But incorporating it within the full true framework – would be as the extreme point of a whole spectrum (or even richer space) of ’partial identifications’: a set of agents’ utilities may be equivalent with respect to certain aspects, where it may then make sense to consider them as a single agent with respect to those aspects. E.g. (as caricatures) a country’s citizens being in agreement on matters of international affairs but not on national ones; or on particularly-pressing current matters (say


\(^{25}\) For cardinal utility functions, equivalence up to affine transformations suffices.

\(^{26}\) This is a contextualized generalization of Team Theory’s basic notion of a Team. As we require the possibility of a transition into a single agent, we opt for different terminology & presentation here.
in times of a necessary war), but not on future ones once the issue is resolved. Or social insects, at the service of their social-genetic group for the most part – bar the chances that they do get to selfishly promote their own particular genome. But even if the space of issues (utilities and actions relevant to them) breaks down nicely (orthogonally) between these aspects and others they may not agree upon, an agent may ‘misbehave’ as part of the unified (though partial) whole, for leverage on matters of disagreement that may be more important to her. Developing a full such framework far exceeds our scope here. But it is important to note that our case of interest (of full utility-identification), is not just an extreme point that is utterly unlikely to appear anywhere in real life. It is an ideal, that actual cases may come close to (or not), perhaps in the particular respects we may be after.

So identical utilities (in their context) determine the single unified agent as an ideal. What does this mean? Could we actually, formally, treat the multitude of agents by the single unified one without changing anything, without losing through this approximation some of the strategic richness, or even distorting its modeling? For this to be the case, further requirements might a priori seem to be in order:

- **Free Communication** (within the union).

  The classical agent is an (idealized) entity that can switch between any actions at its disposal. For the union to achieve the counterpart (in the full product of its agents’ action sets), multiple agents need to be able to change their actions together, in a coordinated fashion. But they all have their own individual information sets, so with limitations on inner communication, coordination problems may arise; the Union as an agent may not be able to switch between some actions.

  A centralized control is the familiar, almost default, solution to managing an agent’s parts-coordination – which of course relies on rich communication to & fro the control center. In this ideal, the 'controlled' parts react *non-strategically* to the orders (which are just information, after all), as the interests are shared, to the point of blurring any agentive distinction within the system. Centralized control may even seem intrinsic to the notion of agent – but of course our work here requires resisting this urge (finding support in the emerging science of decentralized systems).

- **Inner-Rationality**

  For the single agent, the question of normativity, of what it ought to do, can be cleanly separated from what it’s going to do. But things are not as simple for a union. Within one, an agent needs to know that the others will actually *act* according to their utilities – rationally so, to maximize them. This can determine what that agent *ought* to do. And in turn whether the union itself *could* act as one.

- **Common Knowledge** (that the other requirements holds).

  The agents themselves may know that their utilities equalize, that they are in a situation where they are best off acting as one. But they may furthermore need to trust each other to know this, too. And trust each other to trust each other to know this, etc. Common knowledge of the game itself may usually be a reasonable assumption; of inner-rationality – less so. But in any case we will need to keep these in mind, as a most general theory is required for our purposes, treating even genes as agents.
These extra terms are required to ensure that no inner-conflict could arise for which the modelling fusion would fail to capture. They are of course far from trivial, and need not be commonly fulfilled by a union. In fact, a prime example of this work – the genome – bluntly fails the **Free Communication** condition: Mutations in it are standardly take to be independent and rare. And these are the strategy-switches of the *gene*-agents. The *genome* as an agent, then, cannot combine multitudes of such simultaneous switches (and has to rely on series of singles switches that are each evolutionary advantageous, thus leaving vast parts of the genomic space unreachable). Yet the genome still does for the most part act as one\(^{27}\) (for the lifetime of its organism). It’s just that it does so within a much more restricted space of possibilities than would be available without the communication constraint (which constrains the evolutionary search).

Such issues, then, may certainly hinder the union from acting (and from being thought of) as a single agent. But as long the utilities themselves equate, they **coherently determine a stable cooperative optimum of selfish agents as an ideal** (to be sharply distinguished from social welfare optima, which are ordinarily selfishly-unstable, e.g. cooperation in the Prisoner’s Dilemma). So is our premise. And this is what our notion of a union sets out to capture.

On the other hand, for some cases, this ideal may certainly be almost or even entirely attained. What we take to be a single agent may instead be thought of as a perfect union of its different components. Its turning out to be less-than-perfect is indeed what forces us to not consider it a single agent anymore. This is in fact part of that historical revolution in evolutionary theory, towards the single gene’s point of view; towards much smaller agents than had been considered before. Thus, the abstract game-theoretic framework must not commit to a categorical difference between agents and unions, between atomic and non-atomic agents; a priori, they may be of the same type. This should be kept in mind in the face of our adoption and adaptation of Humphrey & Dennett's (1989) metaphor, with states for unions of citizens – which violates that. (Metaphors can only go so far).

This, then, is our notion of a union – as an existing state. What about becoming one?

### 4.2 Uniting

**Section Summary**

The formation of a Union as a (meta-)strategy, with its pros & cons.

In real life, in all its glorious complexity, a group of agents simply happening to share the very same utility function is an utterly rare thing. Yet sometimes, for an agent, it can be a (meta-strategic) **actual choice**, to put one’s self into that position with others, to get into the same boat with them (either instantaneously or through a gradual process). Actions that limit an agent’s available actions limit the possible future realities it can choose from, and its contextualized utility function, restricted to those possibilities, may now equate with other agents’ contextualized utility function.

Keeping to the abstract strategic level of this chapter, let us not attend here to the underlying mechanisms by which becoming a union may be possible (be it the genome or otherwise). A set of agents uniting simply means coming to share their fate *by fiat*. This means that the union’s

---

\(^{27}\) Bar "Selfish DNA".
utility function is some aggregation of its agents’ utilities, who now are (each) best-off pursuing this über-interest (as a manifestation of their original own).

And as we’ve discussed, this meta-strategic choice to unite has (for each agent) both costs and gains:

- **Price of Conformity**

An agent’s giving up on some liberties (available actions), effectively pursuing an interest somewhat different than its original own (the union’s, now), strictly harms an agent – on the first-order level (i.e. not taking into account the game-theoretic hierarchy, the influence in turn on other agents’ considerations. This in particular means that if an agent can fake it, it should).

- **Social Contract Value**

An agent’s Price of Conformity in moving towards the aggregated interest – is the other agents’ gain (as he would come to pursue something closer to their own interest). So looking at it the other way around, each agent gains from the others’ conformity. These choices, however, are theirs, not his, and so may not be up to him. So for rationally-selfish agents, with no mechanism to support the coordination and enforce such a ‘contract’ (as may usually be the case in nature), this value might not enter the considerations whether to unite; It might not drive the dynamics (of the agents’ choices). But it is, in any case, a property of an existing such state of unity as contrasted with the non-united alternative. And in some cases such a mechanism or simply state-of-affairs may be in the background, forcing a choice between uniting simultaneously or not at all, making the Social Contract Value perfectly relevant. And in many other cases yet, the long-term dynamics of multi-stage interactions may de-facto produce a required such mechanism through reciprocation, bestowing upon an agent the gains or losses he brings onto others.

- **Self-Restriction Value**

The high-order value in giving up some options, which can allow the other agents to pursue the union’s interest better (the waiving agent’s included), without putting themselves at risk (of that agent pursing its own, different agenda), and in general trading competitivity for cooperation.

This general game-theoretic point must be stressed: Giving up on some liberties, on the full autonomy, as a single individual’s action (not as some ‘group choice’ to enter into a social contract), may upon first reflection seem to only harm the individual’s position. But higher-

---

28 In general: A. this brings a whole set of complexities and even impossibilities for reasonable aggregations. But our particular applied case will be simple enough to avoid (at least some of) the trouble. B: If the aggregation is not completely enforced by an outside mechanism but can to some extent be determined by the parties, further strategic issues of prior bargaining may arise.
order interaction implications can undermine and even reverse this picture; Self-restriction has its advantages over free cooperation out of temporarily-matching interests. This can be demonstrated through simple examples like this minimalistic game, but the principle is an all-pervasive one, holding even (and even especially) in the most complex of scenarios.

Importantly, unlike the Social Contract value, this one is gained by the same agent making the choice (to take restrictions upon himself). So there is no need for any sort of sophisticated social contract mechanism, only a very simple personal one, a way to commit (or to appear to be committed) to only a subset of its action space. Thus, this value can much more widely be driving agents into unions.

These values and the relations between them depend, of course, on the particulars of the 'game', of the setting. An agent may give up (for the sake of unity) a slightly personally-better selfish option which would have hurt each of the others a lot, while another a personally-favorable one abandoned might hardly benefit the welfare as a whole. The first may then give the agent leverage in bringing the aggregated utility function of the union closer to his own, etc. There is definitely a world of strategic richness here – which we do not get into, and make do with what we need for our basic account of love, and satisfy with raising these issues more explicitly.

It is within such a framework that we can begin to understand and analyze this aspect of agent-interaction – the strategic choice of becoming a unity (including its dynamic dependence on the context in which it is). Agent-unification, then, is the concrete, strong sense we are giving here to the part/whole distinction of the familiar "the whole is greater than the sum of its parts" slogan (which, with "Synergy", are overloaded with related but definitely different meanings, referring to a wide array of phenomena). And we take the Social Contract Value and the Self-Restriction Value together to be the Synergetic Value. The difference between the Synergetic Value and the Price of Conformity determines (at various stages of the 'game') whether a set of agents should unite and whether a Union should disunite – also taking into account the Price of Change. And alternative possible unions, of course.

The game-theoretic notion of synergy set forth is what will ultimately give love its value. But to prepare the ground and help connect the ideas, let us first examine a few simpler cases of such 'acting in unity', 'as one'.

---

29 One particular related notion worth mentioning, from cooperative game theory, is Superadditivity of a payment function \( v \): \( v(S) + v(T) \leq v(S \cup T) \) for disjoint coalitions \( S \) and \( T \). We claim that it is in fact a mathematical principle of cooperation (that can be proved, rather than just an axiom that usually holds) that "in general" \( v(S) + v(T) \ll v(S \cup T) \). But even just giving this meaning is beyond our scope.
4.3 Harmony

Section Summary

A few simpler cases of acting in harmony than love will demonstrate and illuminate certain important aspects of it.

The notion of a Union has put the spotlight on the particular circumstances under which a group of agents should act as one (a normative analysis). When they actually do so, the type of behavior produced is characterized as Harmony.30

Love will (unsurprisingly) turn out to be one rich and complicated instance of harmony (generalizing the genome’s organism-product as harmonious). So let us first present and examine a few simpler instances of harmony.

These will not only help give concrete substance to the highly-abstract approach of this chapter, but set the focus on particular issues that are also at play in love (which, by the time we get to it, will hopefully fit naturally into this framework for the reader).

4.3.1 Ball-Juggling Hands

Imagine someone juggling balls, skillfully. And, fully dressed in black, blending into the dark background, leaving only his hands visible. A ball thrown from one hand is safely caught by the other, which may be moving towards the right position while the ball has barely left the first. And, striving to "put on a show", each throw is as non-trivial as it can be, perhaps trying to maximize speed, height, etc. – under the constraint of keeping it all together, of not making it too hard on the other. But more importantly, together, they set out to perfect some combined, holistic property, which may be a related "maximal number of balls in the air at all times", but also something deeper appealing to aesthetic value rather than just technical mastery. In particular, the roles do not have to be symmetric, but complementary. The quicker or stronger hand may draw the focus in an elaborate trick while the other plays a faithful side-kick to keep it running. They are not in competition; each’s purpose is to serve the higher purpose. If they fail, the fail together; no-one will probably pay attention to which of them is to blame, if such a distinction could even be made.

We would readily and even automatically attribute the actions of them both to a single agent; namely, the person hidden right behind them. But the exhibited harmony is a property of their activity; it is in the visible scene itself. The hands could start moving away from each other, until it became clear that another juggler has stepped in for one of the hands. Could they keep to the same level of harmony? That would depends on the particulars. Two separate minds now managing the performance could, well, on the one hand, direct their separate attentions to more things, allowing for more complicated tricks, or better handling of contingencies, say. But on the other, could not rely on the rich in-brain communication that has previously coordinated the hands’ operations. The harmonious ideal, however, is the same; only how close to it the acting parts can come and in which dimensions skills could be further developed, through training and experience, changes.

---

30 As usual: the term is overloaded with many related but different meanings (mostly generalizing from music, where there don’t seem to be agents involved). We posit our own definition for it here as capturing its familiar generalized interaction sense, without defending this here or getting into it beyond what’s required for our purposes.
The opposite kind of activity would happen if the hands were now tossing balls at each other, entering into rivalry (according to some rules determining a winner). A somewhat different set of skills would now be at the forefront, helping each outtrivial its counter-part (e.g. an ability to quickly re-plan and execute a motoric response). In fact, a single person controlling both arms would have to practice in order to bring about such rivalry behavior convincingly. For agents for which the state of unity is so pervasive, innate and acquired skills are or may become so ingrained within the narrow harmonious subspace of behavioral space, that stepping outside harmony can actually be challenging.

4.3.2 'Togetherness' in Improvisation

(Theatrical) improvisation requires 'putting on a show' and acting well under and responding to a situation not specified in advance (not unlike real life, of course). An extra challenge is added when two improvisers are to cooperate successfully. In particular, in the mirror game, in which two players are asked to mirror each other and create interesting synchronized motion (thereby determining a single criterion of success for the both of them together, putting them 'in the same boat'). One obvious way to manage that is for one to serve as the leader and the other to follow his lead. Could there even be another?

It turns out that expert improvisers can reach a state where it seems (for starters, subjectively) that no particular one is leading the act (with analogous phenomena in music and in sports). It is called 'togetherness' (which "implies unity in which performance of individuals merges into a synchronized whole") or 'being in the zone' ("a state of unselfconscious awareness in which every individual action seems to be the right one and the group works with apparently perfect synchronicity").

Uri Alon & Lior Noy (with others) have set out to explore this phenomenon scientifically. (Noy, et al., 2011) brought the mirror game down to its bare essentials, examining a simplest one-dimensional movement (then extending this work in other papers to further types of measurements and more complex behavior arenas etc.).

What they find is that "moving together is better for expert improvisers: With no designated leader, they reached lower errors in velocity and stopping times, and a wider range of velocities, than when a leader was designated". Some synchronization aspects were too fast to be controlled by visual feedback alone, indicating "that the behavior is not purely reactive but rather has a predictive component", and that "expert improvisers can perform JI [Joint Improvisation], at least part of the time, not by repeatedly switching roles of leader and follower, but by managing to agree on motions together". This, in contrast to the Leader-follower situation which, in particular, exhibited a high-frequency component of jitter (where, "rather than lagging behind the leader, the follower overshoots and undershoots the leader’s motion with a characteristic frequency").
They then turn to offer a simple mathematical model (based on control theory, using reactive and predictive controllers as basic elements) that is able to capture the situation and reflect the basic aspects of behavior demonstrated in their experiment. Their analysis of the model leads them to conclude that it "suggests that the author of the joint motion is the implicit agreement of the two predictors about future motion: The two predictors converge to equal amplitudes in their internal model". And as for this work at large: "These results indicate that people can enter a state of joint improvisation in which both [together] lead the motion, and in which performance is high... The mirror game, involving mutual motor imitation, may tap into a fundamental mechanism of human interaction. Mirroring is a basic form of social communication, thought to be involved in the establishment of parent–baby bonds... and rapport between people".

Of course mirroring is just an extreme, fully symmetric type of joint improvisation and of joint cooperation at large. Whether or not it is a basic form, the "empirical window for exploring the dynamic, cognitive, and physiological aspects of joint improvised action" that they offer here touches upon issues that bear on our extremely-richer case of love just as much. In particular, it brings out the possibility and ability to perform better through giving up on the autonomic agency, reaching synchronization through co-prediction, thereby managing if not bypassing the severe and rigid limitations on communication that is inner-personal.

Hart, et al. (2014) deepens their line of exploration, getting into the (quantitative!) characteristics of the motion to address the question of individuality and its place in togetherness. They "find that each person as a leader does basic strokes of motion with a characteristic signature [in different games], in terms of the shape of their velocity profile between two stopping events. In moments of togetherness both players change their signature to a universal stroke shape [even if they happen to match in their idiosyncratic segment shapes]."
This universal velocity profile... is... symmetric and maximally smooth. Thus, instead of converging to an intermediate motion signature, or having one player dominate, players seem to shift their basic motion signatures to a shape that is altogether different from their individually preferred shapes; the resulting motion may be easier to predict and to agree on. The players then build complex motion by using such smooth elementary strokes”. Analogous considerations are certainly at play in any unification-cooperation that is to be successful, even if the space in our case of love is much, much richer (and much less susceptible to quantification).

4.3.3 Split-Brain

On our journey towards thinking of two brains as together constituting a single self, the structure of the single brain can already bring us pretty darn close: It is divided into two distinct, separate hemispheres (with some functional differentiation, e.g. language primarily localized in the left hemisphere)\(^{31}\). A central a priori difference is of course that they are not completely separate, but rather connected (mainly) through a large strand of neurons (the corpus callosum).

However: For some rare medical purposes, surgeons sometimes cut these nerve bands (partially or fully). Surprisingly, "[such] split-brained patients behave [in general, outside particular experimental settings] in a coordinated, purposeful and consistent manner, despite the independent, parallel, usually different and occasionally conflicting processing of the same information from the environment by the two disconnected hemispheres... Often, split-brained patients are indistinguishable from normal adults"\(^{32,33}\). “In particular, their consciousness of their world and themselves appears to remain as unified as it was prior to the operation”\(^{34}\).

How this could be and what it means (including the particular differences that those patients do exhibit) is being debated extensively in the philosophical literature (beyond the obvious exploration in psychiatry and neurology). But for now we bring this in just as a remark on how close the brain (and the paradigmatic self that it brings about) comes to being composed of two separate parts anyway.

31 Though the lateralization is much less comprehensive than it is popularly taken to be. See https://en.wikipedia.org/wiki/Lateralization_of_brain_function.
33 More minute senses in which the patients do seem to 'split', chiefly under artificial experimental condition, do exist (and are described in these references). These provide further interest, and will be explored in Part II.
34 http://plato.stanford.edu/entries/consciousness-unity/
Chapter Summary

After introducing the necessary general background in Evolutionary Psychology (which stacks the psychological level upon the genetic one), this chapter integrates all the components of the previous chapters into a unified framework.

Sexual reproduction and parental investment are brought into the evolutionary-psychology picture, in preparation for the application of the framework to them.

5.1 Background: Evolutionary Psychology

Section Summary

A brief introduction to the field and way of thinking, and how it relates to this (non-representative) work in particular.

The biological world supplies for great many issues whose current understanding demonstrates the principles of evolution, which in turn provide for a unifying framework that integrates them all. At its core, Evolutionary Psychology is all about approaching psychology through accepting the brain as just another organ, answering to the very same principles, shaped by them, with the hopeful promise of bringing the great many various aspects of its workings under a unifying framework. It seeks to integrate psychology into modern scientific evolutionary-biological understanding, accepting "that much of human behavior is the output of psychological adaptations that evolved to solve recurrent problems".

The field’s origins go all the way back to Darwin, but it has flourished in the past few decades. Just as with the Gene’s Eye View (and in continuation of it), there is a whole world-view behind this. Familiarity with this way of thinking, indeed this view of life, does not come easy, and cannot be truly brought here into our narrative through covering of some main points — though it is necessary for coming to accept and be convinced by our narrative, and filling in for the many quandaries that are bound trouble the reader, about how this is all supposed to work. Two recommendations here are the rather popular (but high-quality) Pinker (1997), or the academic field’s foundational text – Barkow, et al. (1992) (with quite a few chapters on topics most relevant to ours). We will, however, make do without assuming acquaintance (but without boring the acquainted too much), and bring in only a few particularly-relevant points.

One central challenge when approaching the brain over just about everything else in the biological world is how far we are from understanding its workings, and how these come about from the underlying genetic basis. This sets back the possibility of doing equivalent current-quality science in this realm. But in principle, the evolutionary perspective can still be valuable and unifying, just as it had been for biology before the discovery of the DNA or even the acceptance of Mendelian inheritance. In particular, illuminating the brain’s teleological function (what’s it for), exploring its product – the mind, and its product – behavior, is still possible at the more general, principled level. Evolution provides for the normative, rational background

35 https://en.wikipedia.org/wiki/Evolutionary_psychology
36 Which also presents (Frank, 1988)’s "reverse-engineer[ing of] mad love".
37 Or the more updated (Buss, 2005), which furthermore has a new two-volume edition that just came out (2015-16).
against and within which actual (perhaps contingent) facts about brain-mind-behavior are to be understood. Psychological predictions can be derived from an evolutionary analysis, which in return can be re-evaluated and refined when such predictions don’t turn out to be the case.

As stated so far, very little about evolutionary psychology should have been at all controversial – which is in fact as far from being the case as any scientific field would have it. Part of the fuss stems from a priori explosiveness – from the subject’s direct implications for matters close to everyone’s heart (rather than just scientific experts), bringing in ethical and other political emotionalities. But there are also deeper methodological issues that bear also on our own work here.

The connection from genes (through brain) to behavior is very much unaccounted-for (in details). The high-level nature of the phenomena explored thus makes for an inherent obstacle, in telling apart true adaptations or exaptations from by-products or random variation – from cultural-environmental influences. Finding cultural universals, such as language (its existence but furthermore an underlying shared structure of Universal Grammar), is a common starting point for evolutionary analyses. And in the other direction, evolutionary thinking putting a spotlight on certain issues as central and universal to human existence, can guide a search for cultural universals – at times in the face of existing beliefs. Love itself, in fact, is one such topic, which at certain times has commonly been attributed to be "an invention of western culture" (Sternberg, et al., 2006; Campbell, et al., 2005). Our own focus throughout this work and ahead, on very basic biological givens, will of course suggest an utterly different picture (as in fact turns out to be the case, after the conceptual presumptions that prevented or distorted empirical examinations gave way).

A central common methodology of the field that goes beyond the above uncontroversial generals, is the following rather concrete analogy that is made between the workings of the mind and of everything else in these evolutionary explorations. A body part can be conceptually set apart and explored in isolation, in terms of its functionality (e.g. the eye and what it is for), within the more general structure in which it interacts with the rest of the body. Likewise, many quite-specific behaviors can be set apart and evolutionarily analyzed on their own. For example: in the contrived scenario of the Wason selection task people tend to fail at utterly basic deductive reasoning, while doing perfectly fine in a logically-equivalent real-life scenario of detecting violations of drinking age-restrictions. This has brought about postulations of "a specialized cheater-detection module". (Barkow, et al., 1992 pp. 163–228)

This approach, of setting apart and focusing on specific types of behavior, thinking of them as if they could independently-enough be selected for or against, is intimately tied to the thesis of massive modularity. This is an assumption about the brain containing many "functionally specialized cognitive systems that are domain-specific and may also contain innate knowledge about the class of information processed". Although very much under dispute in cognitive

38 For a succinct summary of the common general criticisms and some responses to them, see https://en.wikipedia.org/wiki/Criticism_of_evolutionary_psychology

39 See categorization at https://en.wikipedia.org/wiki/Evolutionary_psychology#Products_of_evolution:_adaptations.2C_exaptations.2C_byproducts.2C_and_random_variation

40 See http://plato.stanford.edu/entries/modularity-mind/#MasMod

science, it is thus by and large endorsed in evolutionary psychology for its methodological benefits (allowing to take an evolutionary look at specific-enough issues)\(^\text{42}\).

This is mentioned here to emphasize that this is not the route our own work takes. This is not to rule out any tailor-suited mechanisms being involved, supporting the phenomenon and even making it possible. But a unification of selves – the whole selves – goes beyond any particular domain-specific issue – as much as anything can. Our appeal to evolutionary psychology will consist only in relating the interests of the genes as a whole (as a complete genome or two) to the interests of their psychological edifice as a whole.

And so this last distinction, the bearer of common misconceptions, should be clarified. The genes’ interest are what evolution determines directly; the normativity it brings in (where a gene following its interest is that which would tend to be found in nature). The psychological interests are a property of the brain they (in fact the cogni-genes) construct. Although hard to define as well as explore (in a different way than the genes’ interests are), these are the ones familiar from our normal lives, having to do with pain, pleasure, and the way we actually behave, the things we in fact pursue. These approximate\(^\text{43}\) the interests of their genetic constructors, indirectly – and all the usual caveats apply. A changed environment, for example, can get in the way of identification of the two levels of interests. A sex drive, which has been implanted to set up the psychological interests according to the genes’ own (a.k.a. kids), goes astray in a changed, modern environment, that is now contaminated\(^\text{44}\) with contraceptives. And given the immense changes modern life brings upon humans more generally, this necessitates delicate care.

5.2 Tying it all up

Section Summary

The synthesis of all the levels and principles introduced previously, into what is our full framework.

We are now in a position to bring in everything together, synthesizing the components, the principle that connects them, and the principles that govern them – into what is the full framework we put forth.

Its substance are the two levels – the genetic level at its base, and the psychological level on top of it, as constructed by it. The first level’s agents are the genes, while the latter’s are some ‘citizens’ according to Humphrey & Dennett’s metaphor. (It is not essential that such mysterious person-like entities could literally be found in the mind, just the principles by which

\(^\text{42}\) The methodology could to some extent be defended even if massive modularity is wrong, considering that general-purpose mechanisms, though more abstract, should, too, have rich-enough representations to act appropriately in particular domains, as if they instantiate domain-specific mechanisms. But this is taking us off topic.

\(^\text{43}\) This is known as the Proximate vs. Ultimate Causes distinction. [https://en.wikipedia.org/wiki/Proximate_and_ultimate_causation#In_biology](https://en.wikipedia.org/wiki/Proximate_and_ultimate_causation#In_biology)

\(^\text{44}\) Well, from the genes’ point of view.

Circumventing ‘blind’, basic instincts or their results for the sake of better consciously-planned conduct can certainly make sense and is quite common. And alternative drives for reproduction may integrate better with that, for the sake of better offspring-number regulation – which is quite common in nature (e.g. "The trade-off between number and size of offspring in humans and other primates" / Walker, Gurven, Burger, Hamilton, Proc. R. Soc. B, 2008). For example, we seem to be predisposed to being fond of children (indeed, the evolutionary forecast in this contraception-contaminated environment would suggest that such dispositions would take a front role, be selected-for). However, current birth-rates in most rich countries strongly suggest that behavior in the face of available contraceptives is mostly a serious case of maladaptation.
they bring out a self and reflect and act according to gene-dictated interests, so bear with us. And it is certainly not required nor even reasonable to expect that each agent of the base level map onto (i.e. be in charge of creating) an agent of the top level. Each level brings with it its own sense of a union: a coalition of genes (the genome being the prime exemplar), and a 'state' of its citizens (which brings about a (conscious) self as head-of-state, the classical self being the paradigmatic one). Each one allowing for other unions than the paradigmatic ones that, also being a multitude of agents working together harmoniously, are not essentially different.

Evolution endows the base genetic level with normativity – the utilities of the genes. Evolutionary psychology entails that (whatever the bio-neuro-logical details,) these utilities are transferred (approximately) to the utilities of the top psychological level, its phenotypic manifestation: a union of genes, as a whole, is to be represented by a union of psychological 'citizens' (a 'state'), as a whole. According to the interests (that change with circumstances and choices), each level – in principle – allows for the dynamics of splitting unions and merging into bigger ones, in accordance with the strategic tradeoff (of the Price of Conformity vs. the Synergetic Value, plus the cost of change). As these are abstract strategic principles, pertaining to both levels correspondingly, the citizens would (in particular) choose to merge or split precisely when the genes would like to have them do that.

Such dynamics are possible in principle, but not necessarily in practice. A single genome’s genes are biologically confined into a shared unity, restricted from splitting. There is no similar theoretical restriction on the possibility of merging. But the organism’s body itself is a rigid, rigid,

---

45 Recall that:
A. There is no single-agent mapping between the levels.
B. A hefty apparatus is required to construct an interest-representing cognitive system, of the sort that could participate in the dynamics we are after.
And that agents in a union take after themselves by taking after them all. So each gene has an interest to do its part in creating the psychology-level agents as in a (corresponding) union.
46 Except for the usual approximation of the genes' interests by the proximate mechanism. The (perhaps imperfect) correspondence of interests between the levels transfers to a (correspondingly-precise) correspondence of split/merge preferences.
fixed product, that cannot be physically merged with another. The underlying non-cognitive genes are not only a priori stuck with each other – but with no other. Not at all so for the cogni-genes; not at all so for a decentralized cognitive system they may construct. Information is an abstract entity, that transcends physical boundaries. This is where the self as head of state framework comes in, fitting in naturally. Its authors’ dynamic theory of selfhood-formation may allow not only for new elections as needed (e.g. childhood personality formation) – but for elections to take place within and for a larger, newly-joined political entity – selecting a new head-of-now-united-states.

In what context, then, might a gene find itself in such commonality of interests with genes beyond its own genome, so as to prefer forming this new, enlarged, cognitive unity? At this stage, the answer should hopefully be evident: sexual reproduction.

5.3 Background: Mating Behavior and Parental Investment

"Survival of the fittest" makes sense as a slogan about the species level; indicating which ones are to be found. Unfortunately, it can be misleading, being taken to be about the individual organism; putting the emphasis on its survival. Which is not the heart of the evolutionary process. Rather, maximizing the number of copies (of the genes) is – and survival is just means to an end (so to speak). This means that psychology itself is actually shaped for the sake of reproduction, the ultimate motif, running deeper than even the will to live\(^{47}\). And the modern view, centered on the agent as a selfish one, puts the spotlight on the strategic complexities that are inherent to it all (as opposed to some harmonious ideal of mating / parenting that necessarily-imperfect beings simply fail to reach):

"Reproduction is always costly... Individuals are limited in the degree to which they can devote time and resources to producing and raising their young, and such expenditure may also be detrimental to their future condition, survival and further reproductive output. Parental investment is any parental expenditure (time, energy etc.) that benefits one offspring at a cost to parents' ability to invest in other components of fitness. Components of fitness include the well-being of existing offspring, parents' future reproduction, and inclusive fitness through aid to kin... Overall, parents are selected to maximize the difference between the benefits and the costs, and parental care will be likely to evolve when the benefits exceed the costs"\(^{48}\). In particular, an individual organism’s interests are split between itself and an offspring from the very moment of its creation.

To all that, sexual reproduction adds a world of further complexities. For example, when a single parent’s investment may be sufficient, each side may prefer leaving the other stuck to take care of the offspring, while going away to make new ones. Furthermore, "The size differences of gametes... males (small gametes—sperm) and females (large gametes—ova)... typically results in different levels of parental investment between the sexes, with females initially investing more... For male mammals, which have a relatively high maximal potential reproduction rate, sexual selection leads to adaptations that help them compete for females. For female mammals, with a relatively low maximal potential reproduction rate, sexual selection leads to

---

\(^{47}\) Incidentally undermining the very foundations of common ethical, "humanistic", individual-centered world-views. But that’s a topic for a very different paper.

\(^{48}\) [https://en.wikipedia.org/wiki/Evolutionary_psychology#Parenting](https://en.wikipedia.org/wiki/Evolutionary_psychology#Parenting)
choosiness, which helps females select higher quality mates\textsuperscript{49}. (It should be noted that for love, (at least by our approach,) these differences matter much less; most examples and analyses will be independent of this generally-fundamental asymmetry).

The centrality of reproduction from the evolutionary point of view, has made its implications both a central focus of evolutionary psychology\textsuperscript{50} and perhaps its biggest and most-paradigmatic success. In terms of predictive powers\textsuperscript{51}, and the ability to give theoretical unity to many cross-species and cross-cultural variations. (One glaring current inadequacy – though not an outright contradiction – is in explaining homosexuality\textsuperscript{52}). However, due to theoretical and obvious methodological reasons, the field tends to focus quite exclusively on the simplest, most observable and quantifiable of behaviors. Much less has been done in the very much related but most high-level of evolutionary psychology frontiers that is love, empirical access being mostly through subjective reports for humans\textsuperscript{53}, and not-at-all for other species. Limitations that certainly come as no surprise by our own approach, given the all-embracing mental and behavioral nature of the phenomenon. See (Campbell, et al., 2005) for an overview (and as a more thorough introduction than this section).

\textsuperscript{49} \url{https://en.wikipedia.org/wiki/Evolutionary_psychology#Mating}
\textsuperscript{50} In fact, quite a few of the papers of the field’s foundational text (Barkow, et al., 1992), are about these issues. Which of course provide for a much better introduction than such Wikipedia excerpts. Particularly recommended is Wilson, et al. (1992).
\textsuperscript{51} Just a random example: "Evolutionary thinking led to the discovery of the most important risk factor for child homicide – the presence of a stepparent." (Sternberg, et al., 2006)
\textsuperscript{52} There are some partial possible explanations and illuminative findings (e.g. \url{https://en.wikipedia.org/wiki/Fraternal_birth_order_and_male_sexual_orientation}), but the problem (theoretically speaking) for the most part remains open.
\textsuperscript{53} See Section Error! Reference source not found.
6 Love

Chapter Summary

Our general framework is applied to the context of sexual reproduction (with parental investment), deriving (under certain specific conditions) a result most appropriately termed "love". The strategic space begins to unravel, and some basic connections are made to the phenomenon as familiar to us as living human beings.

6.1 The Heart of Love

Section Summary

Love does not consist in gazing at each other, but in looking outward together in the same direction. – Antoine de Saint Exupéry

Love is all around us. We know it directly through personal experience, and through the many people that are our social environment. Such experiences are embedded into our folk-psychology – the non-scientific collected data and psychological theorizing of the laypeople. This is the environment in which that principal metaphor of soul unification thrives. Some folk-psychology may certainly be incompatible with, rather than just partial to, professional, 'scientific' psychology. Yet it isn't simply inferior to the latter, but often possesses vague truths science has yet to reach, crude insights that could help direct it. We've taken our cue from it for the leitmotif that runs through this work. And now, we finally reach the phenomenon as a conscious one. Let us unveil basic interactions with our framework, and hopefully shed some new light on the folk-psychological conception in return.

The organism, recall, represents its genes’ projected estimation as to the best way to maximize the number of copies of themselves. Their biological product reflects this, and their psychological product in particular – just as much. Now, sexual reproduction means that for the genes of a genome, the fate of each and every copy they produce is – necessarily – perfectly intertwined with the fate of another’s (with the fate of a copy of another genome’s genes). In other words, the utilities of the two genomes relative to a single offspring coincide completely. And more generally, this applies to a single spawning (with the same single mate). For a clear, simplest picture of what this can entail in light of our framework, let us introduce the following ultimate thought-experiment –

Single-Copulation Species

Consider an organism who (as a property of its species) is restricted to copulating only one single time throughout its life though living to see another day54. The choice of a mating partner may be the single most important decision of its life. Once the deed is done, there’s no going back. All that’s left is to take care after its offspring the best that it can – in complete, harmonious coordination with its spouse. Its own ‘personal’ autonomous mental self is of no use anymore, and can dissolve just like the salmon’s body right after it has gone upstream and spawned

54 This might not be evolutionarily viable, as it seems better in this case to 'put in everything' into the spawning itself at the cost not living past it. At least that is the general rule. See
https://en.wikipedia.org/wiki/Semelparity_and_iteroparity#Semelparity
(“fertilizing the stream and creating a nutrient-rich environment for the new infant salmon”\cite{55}). Dissolve, and fuse, together with the chained-up spouse, into a new mental duo-self instead\cite{56,57}. For them, to make love is to literally make love. ’Till death do them part.

What kind of an entity is this, though? Tight action requires a tight cognitive connection, it would seem. So how deep can the cognitive integration go? For example, does the male actually experience the female’s physical pain? In principle, it would seem that he should; her arm is as important to the unity from either side of it. But that pain serves her to notice harm and avoid further damage. When something happens to her, she is the one who’s always physically there to actually do something about it (the couple always sticking together would be profoundly non-optimal, for example preventing the sexual division of labor in hunter-gatherer societies). She ordinarily has the means to respond the best (and certainly the quickest), even if the damage is – in the important sense – to the unity as a whole. So the pain has a local, particular-body function; it indeed should still be chiefly inflicted upon her, even if her spouse truly does care about it just as much. The point here is a general one:

It is fine that much of each cognitive side is still dedicated to its own body (and to itself). The duo cognitive system may be united in our interests-centered sense, though it is decentralized, lacking a neural, information-rich connection – and aiming for the optimum (duo) functioning as such. A limited interface of communication (through talk etc.) necessitates that coordination be rather high-level: The goals need to be in long-term or abstract terms (“Let me worry about keeping the place clean”). And (as with the smooth elementary strokes that are easier to predict and to agree on of section 4.3.2 - ‘Togetherness’ in improvisation) the operations need to fall into general styles (“At the risk of sometimes hurting people, I’ll be witty, but keep me in check so that I don’t go too far, and try to fix and ease the situation when I do”). Given each side’s perception of a same situation, coordination can go a long way even with very little communication: With no inner-conflict between them, their predictors of each other can converge, supporting their joint action (as in section 4.3.2). But as for handling the finer contingencies that each side – body and mind – is bound to stumble upon as life happens, they are technically on their own. Physical pain was but one rather-simple example. Such matters of Free Communication are at play at large, determining a considerable extent of (localized) personal autonomy that keeps hold even within fully-united, ideal love.\cite{58} (All this is completely analogous to how a single brain’s cognitive modules can be rather autonomous, the restricted interface between them being high-level, esp. as in section 4.3.3 - Split-Brain. The general point,

\begin{flushleft}
\footnotesize
\textbf{55} Random internet source.
\textbf{56} - "Those who love have, so to speak, pawned a part of their narcissism." –Sigmund Freud
\textbf{57} - "the assent by which we consider ourselves henceforth as joined with what we love in such a manner that we imagine a whole, of which we take ourselves to be only one part, and the thing loved to be the other" –Descartes
\textbf{58} - "If I yield myself completely to another and obtain the person of the other in return, I win myself back... In this way the two persons become a unity of will... Thus sexuality leads to a union of human beings, and in that union alone its exercise is possible." –Kant
\end{flushleft}

As usual, we do not care for interpretations of historical figures and their correctness.

\cite{57} Recall (Humphrey, 2000) as discussed in section 3.3, which would seem to suggest even the growing of a united consciousness.

\cite{58} We will work out these implementation details of the cognitive architecture ourself in Part II, but the reader is strongly encouraged to read (Wright, et al., 1996) – through these lens, recasting it (when necessary) in our terms.

41
as always, is that there is no major principled uniqueness to the particular type of unity that is love. So each side is autonomous to an extent. But the autonomy is tactical, not strategic.

### 6.2 Commitment

**Section Summary**

The relation of love to commitment is explicated, strategically and phenomenologically.

Love is not love Which alters when it alteration finds
– *Shakespeare*

From the outset, at the strategic level, to fall in love is to commit: to give up on the ability to pursue the single genome’s interests on their own regardless of how much these may diverge from those of the spouse’s genome (and hence from those of the combined duo-genome). Some of the familiar conceptualizations of love certainly give commitment a central or even constitutive role (Sternberg, 1989). Our explication of game-theoretic unity allows us to give commitment this scientifically-motivated, clearer, concrete meaning. For example: Should the perfectly loving imaginary creature (a male, say) be prepared to die for his mate? Maybe yes (e.g. if she’s pregnant and he will probably soon be dead anyway), maybe no (e.g. if what the offspring will require most is physical protection, from their stronger father) – but there is a well-defined answer in principle (that will maximize their offspring’s prospects). The manifestation of his commitment as a will to act (in accord to it) can be mediated through (reflecting his mate’s) pain & pleasure. In general, then, how deep should his empathy run? As we’ve seen, there’s no need for him to actually experience her physical pain. For other types of pain (e.g. losing a parent?) and other experiences – perhaps more ‘mirroring’ is appropriate. It all depends on the how localized the *function* of the phenomenon is (to the particular body or brain). The commitment is at the most abstract level; its actual various manifestations depend on the whole of particulars of the setting. All other senses commonly attributed to commitment do not stand at its core, but are derivative properties of unity.

The value of love (which we denote by $L$) is how much there is to be gained for the sake of the shared offspring by the parents uniting into a single agent, their two cognitive systems into this gestaltist whole. The imaginary single-copulation species determines the *ideal* of love. In this extreme, artificial case, the value of love does not matter – there’s simply nothing to lose. But we shouldn’t expect to find such a situation in nature, as it indeed does matter, a great deal – for humans too, of course. (And humans shall be our main focus from here on, aided by our intimate acquaintance with them, at the psychological level. Tying it all to behavior as observable in other animals too, e.g. monogamy, is deferred to part II).

**Don Juan**

At the other side of the spectrum, where a male may have great many opportunities to impregnate a diversity of females, each quite capable of bringing her young into successful adulthood by herself, the situation can be different. For him, the value of falling in love with one woman for the sake of their shared children may be far outweighed by the cost to his offspring-maximizing freedom. As his lady prefers having him around, focused on promoting

---

59 Which, as usual for evolutionary psychology, is devoid of any ethical contents. One may still be an ethical realist, taking life to be about something other than the evolutionary imperative.

60 This is the only point in this document where the gender-asymmetry matters.
only her children, rather than the potentially larger group of all of his, so would the duo-self prefer, to some extent. This means him not being able to 'screw around' as much. (Having their own set of interests, mistresses may require resources, or even try to steal his heart – in direct opposition to the former union's existential interest). And worse (probably), it means not being able to simply run off on her and their new-born, in order to have it possibly manage without his support, while going on to produce more (that he may opt to stick around with) elsewhere, with a new mate. All this is not to say that womanizers don’t fall in love, of course. Exploiting an unlikely contingency of turning up extremely attractive (in the general rather than just the physical sense) may not be supported by the genetically determined innate makeup (as guided by expectancy). Nor does it mean that womanizers shouldn't fall in love. Many factors factor in: the effort required for such a lifestyle; the single-mom’s ability to raise a child on her own; etc.

Frank (1988) puts all of the weight of love on the issue of mate-choice optimization – satisficing (making do) with an approximately best available choice and committing to it, in the face of future possible better ones. By our own account this is just part of the story; love is first and foremost about the internal couple-dynamics. The forgoing of options in uniting means that the unity has to hold even in the face of only lesser spousal options (even if that spouse is "the best he/she can do"). A wife might be better off depleting the shared account (which was created in trust) and leaving with it to some somewhat-lesser mate. Or, more fundamentally, a husband might simply run off on her (as just discussed). Unity – strategically and mentally – runs deep through a couple’s whole interaction; this is the source of the synergetic value. Still, this issue of mate-upgrade prevention is a valid and eminent one, certainly interacting with love-quaque-commitment.

Heartache

Section 1.1.2 introduced John, who’s recently been dumped by Jane, unexpectedly. And who, despite seemingly being perfectly fine in all 'objective' respects, for all future prospects, is utterly depressed, feeling and behaving as though there’s no point going on living. This actually does make perfect sense, when we realize that we are not really dealing with John per se here; not as an autonomous cognitive entity. But rather, with a part of the former John-Jane being. A being which may now be quite literally (evolutionarily) deceased, having no prospects for future generations. Or perhaps dying, with nothing better to do for John than hold on to the chances of getting back together, obsessively, disproportionally, however slim they may be. We might be tempted to wonder why it should be so hard for him to 'just get over it'. But remember: that’s exactly the point of falling in love. Giving up the autonomy, allowing the other side to trust the first to act within the duo’s best interest rather than its own. This means that splitting should be costly, not a readily-available option. From John’s point of view, Jane leaving him is in the important (evolutionary) sense actually equivalent to her dying on him61 (before there are any existing kids). His not being ok after such a thing would happen drives him to prevent such a thing (either one) from happening in the first place, even at a considerable cost or risk to himself62. The primordial risk John took upon himself was in entering this union. And the fact that it didn’t work out, in itself, of course does not in any way mean he was being irrational.

61 Breakups are indeed compared in the clinical psychology literature to death in terms of the mourning they can bring. But the evolutionary perspective gives a much clearer theoretical reason to put these in the same category.

62 Here, as usual, no sex-roles are meant for.
Still, one could wonder: if love is this commitment contract, why not include a clause stating that in this particular kind of case, with Jane leaving out of her own autonomous volition, John should just be freed of his commitment? What’s in it for her that he should now suffer?

Well, the commitment is not a contract. There is no external mechanism to enforce just any particular terms the sides may want to agree upon. The commitment is purely in each’s mental state, and in the many components of the interaction, supervening over their whole shared life as embedded in their environment (e.g. one taking care of all the bills and the other not having to worry about it). John, while in the union, was going to suffer if thrown out of it – and that’s precisely how Jane needed for him to be, when she was in the union, even if now she’s in a different place and not gaining from his pain (or even being harmed by it in some way too). His whole state cannot simply be disunited at will. There is and could be no single bit of his mental state allowing Jane to simply notify him that she’s fine with him not suffering from breaking their commitment on his part too ("I’m not in love with you anymore, it’s done, move on"), and through it free him of his commitment. If there were, then as there is no external enforcing mechanism, he could simply act as if that signal has been received and free himself from the commitment at any time, out of his own volition. Unity is a deep property, not a patchy add-on feature.

6.3 Person-Fusion

The synergetic value – love’s value – results from the many different changes that the unity allows for – and that constitute the unity. The self-restriction value can determine a long-term dynamics of willingly giving up more and more liberties, gradually going further and further into the union ("growing on each other"). Meanwhile, the social contract value is much more all-or-nothing, the ‘contract’ being enforced by the offspring who is being created, through the drastic utilities alteration that it brings. (Though this value too is somewhat gradual – impregnation being a random process, and partial – given the possibility for other past and future kids). Both values together rely on the many changes in each side’s life and mind that are reasonable to make only given that the other is undergoing a complementary waiving of options. This is not true for a single act done for the sake its self-restriction value. But within an iterated dynamic of many such acts, each may depend on the updated, more-united state (the other side having already given up on some liberties of their own), without which that liberty should not have been waived.

How to define and conceptualize love (in terms of its types, or components) is a central debate in the relevant psychological literature (Sternberg, et al., 2006). Many different approaches are contending, but there is a fundamental theme arising from it all (including even through quantitative meta-analyses of the different approaches): The division into two basic types – Compassionate love vs. Passionate love. "Compassionate love is defined as.. a friendly affection and deep attachment to someone.. This kind of love is characterized by caring, trust, honesty, respect, and the like.. It can be experienced for a number of significant people in one’s life (e.g.,

63 Thanks to Elchanan Ben-Porat for doing so.

64 A Cryptographic system could perhaps allow her to have a key to his system that he himself does not poses. But such mechanisms don’t tend to show up in nature.
close friends, family members, romantic partners). Passionate love, in contrast, is defined as a state of intense longing for union. This kind of love is characterized by emotional extremes, physiological arousal, and sexual attraction. Connecting the emotional to the strategic, it seems possible to relate compassionate love to the self-restriction value and passionate love to the social contract value (despite their vagueness). In particular, as we’ve mentioned, the dynamics driven by self-restriction can be quite gradual, and necessitate no binding-mechanism (of the sort that sexual reproduction brings). Thus, it is a major force in forming close non-romantic relationships – while only a subordinate force in romantic ones (where it is dwarfed by the genetic stakes, the potential for extreme interests-binding). But we leave the development of this connection and the interaction between the components in romantic love to part II.

The dynamics of falling in and out of love are another topic that surely deserves explicit attention, much more than we will give it here. For example, how did Jane (John’s heart breaker) come to disunite on her part, possibly without John realizing this? Shouldn’t it have gradually produced disharmony, that, as a property of the interaction, could be detected by John? Could deteriorating harmony be told apart from alterations that love ought to withstand? Etc. But let us begin by trying to think about the dynamics of love in terms of personality change, as our framework suggests.

**Personality & life dynamics**

As Humphrey & Dennett note, personality is a choice, solving an optimization problem of how best to achieve the (genes’) goals, with the organism’s given innate capabilities, within the environment it finds itself in. Being relatively smart, for example, may direct one to particular ‘better-paying’ sub-niches that make better use of it; and having no option to enter such a sub-niche (or merely thinking one has no such option) may on the other hand sway him away. Personality builds up throughout the individual’s life, changing slowly and gradually for the most part, slowing down from childhood on as his estimation refines.

Yet individual maturity is not the final stage – but creating offspring and taking after them perhaps, possibly through a union with the spouse. The individual personality in form, then, should be prepared for the eventuality of deeper bonding. And still, when a particular such bond forms, it may constitute a radical departure from that gradual course of the self throughout its life, as it (relatively) abruptly forms a part of a new duo-self in its place (possibly quite different from previous ones he may have formed, too).

To be sure, this drama isn’t necessary: A standard rigid socket may take in many different appliances of a fitting plug. A husband (to opt for the classic sexist caricature) may be all about making money, in need of a wife who’s all about caring for children. As long as there are enough of each type, aiming at this kind of personality-choice may be a reasonable bet. A man can focus on becoming the best he can in the relevant respects (e.g. the ‘manly’ qualities common in society), so as to improve his value in these dimensions, for its own sake as well as for attracting the best women in the completing dimension. (And likewise for a woman). Various alternative niches can thus coexist (e.g. ‘feminist’ couples), as long as the preordained completing potential spouses don’t become too scarce (making the task of finding one too hard / the degree of necessary compromise too high).

---

65 From Fehr’s paper in (Sternberg, et al., 2006), which offers a great brief review of the topic.
Conceivably, then, rigidity could work. But perfect rigidity is in general suboptimal. It entails a narrow matching-interface between the two parts (with substantial personal autonomy within the relationship – in the subspace of the united being each side is in charge of). A shallow integration which results in a union of lower synergetic value. It leaves no room for the extra optimization that is possible given the two particular parts with all their unique traits and capabilities, on the way towards unity. It forgoes the unique professionalizations that this particular unity may be able to reach. The optimum (partial to the unconstrained ideal) would involve a deeper integration. A substantial reformation of the two states now united, appropriately headed by a duo-self that is much more than just the gluing together of the original selves. And a deeper integration means a deeper bond. This, we posit, is a central component of the commitment mechanism. This – as we’re in the business of substantiating clichés here – explains "love’s ability to change a person".

To briefly detour into an extreme, pathological instance: for Humphrey & Dennett, multiple personality disorder usually having to do with early childhood sexual abuse is but a curious fact (they tie the disorder to extreme emotional stress, which theoretically may be caused by completely different things – which however do not tend to cause the disorder). But given our connecting of sex to love to selfhood & personality-formation, this empirically-found connection (between the disorder and its alleged cause) seems far less arbitrary, if still not at all explained. But let us leave this kind of speculations aside, and go back to much more common, well, states of affairs.

Falling for someone

We’ve discussed at some length the symmetry-breaking of (e.g. John) being left alone in the (deceased) union. The complementary case is that of entering a (non-born) union alone; Brian falling in love with Brianna, one-sidedly. This, to state the obvious, is an extremely risky, arguably desperate strategy (putting Brian in a situation basically akin to John’s). Its only value can be in boosting the chance for that potential union with Brianna to actually become a reality. How would that be a seduction?

A process of becoming a couple, reforming into one, is an experiment, that may be lengthy, pricy, and at high risk of ultimately failing. Typically, Brianna too may get hurt (e.g. from someone who will turn out to just be using her for sex). Brian, by making a significant leap up-front, on his own, reforming himself and his life to fit her, takes much of the cost and risk upon himself66. He may, for example, change his location, to keep near her and hopefully get her. Or support her financially or buy her gifts – money which is not lost only if they come to unite (signaling how much she means to him, how much he values her relative to himself). Or he may get into poetry (which fills her world) to gain a richer common ground on which they could communicate more profoundly. Or constantly think about her and the things she says, to learn to see the world through her eyes, to strengthen the cognitive connection (his predictor in unity). Reforming to suit her specifically (instead of improving himself in general) makes his commitment to her a credible signal. As he is already hers, Briana can be fairly (or at least relatively) certain that if she’d want it to work, it will. But more than that, it can convince her that she should want it to work:

---

66 Though not all of it, as Brian cannot do his whole part in the reformation a priori. The informational interface between their brains is very limited (i.e. he doesn’t really know her completely), and so future interaction necessarily has to work out further details – and may still fail.
To begin with, the (proxies for) qualities that Brian displays (money, academic degrees, friends,...), the ones observable by Brianna, are of a general, impersonal sort (which he has been supposedly maximizing throughout his life). She can only extrapolate from these the symbiotic value that could emerge between them. But Brian has extra information that she does not possess. Namely, a much better acquaintance with himself: with his deeper qualities (rather than just somewhat-arbitrarily-chosen realizations of them); with everything that he could be, could still become. This means that his estimation that he and Brianna could form a highly-synergetic, worthy couple (for her too!) – an estimation that she apparently does not share – may in fact be justified (though may also not be, as he suffers from the symmetrically partial information). And so, by becoming his side of their union, Brian can bring her to see what really matters: not his qualities in general, but his qualities for her as a soul-mate; their synergetic potential. This is the risk he may have to take in order to show her that they indeed should be one.

**Being 'special'**

To address one last phenomenological expression of love: Is the appreciation of one’s spouse all about their objective qualities? In the course of falling in love, the other individual actually does become unique "in the eyes" of the first. Finding the spouse to be 'perfect' in spite of objective formal qualities and cold-hearted material calculations (or even re-adjusting the view on such matters in order to bring out the spouse as perfect) is not a matter of purposeful self-deception. It is not just an ad-hoc (rational) tool for commitment to irrationality in the face of possible future upgrades (as Frank (1988) would have it). It is at most a case of consciously-verbally making up reasons for that "of which reason knows nothing"67. And more than that, it may actually carry an objective sense, of the spouse being quite perfect as a completion of the first individual (following a reshaping of the two personalities), in a world where people aren’t just ranked on a one-dimensional quality scale. (This does not dictate any sort of pre-destined 'soul-mates', given the personality-plasticity).

There exists research on matching, which tends to demonstrate the surprisingly-large extent to which simple quantifiable parameters can determine quite a lot in people’s choice of mate (and even more simplistic predictive models in biology of animals lining up according to size and the likes). But surely there is something to people’s perception, that the challenge of finding a match goes much deeper, having to do with the richness of the personalities involved. We hope to have convinced that this may also be explorable at least in a more mathematical tone (than folk-psychology), even though what we’ve offered is hardly even the beginning of a proper analysis. (Also leaving so much untouched: To what extent is the level of personal autonomy a general pre-unity trait of the individual, or how does it depend on the particular pair? To what extent can a personality change survive disuniting? And many more).

---

67 "The heart has its reasons, of which reason knows nothing." – Blaise Pascal
6.4 Other Influencing Parameters

We have denoted the value of love\textsuperscript{68} by \(L\) to set it apart and emphasize the centrality of this value. But this would be a first step towards putting this theory to some more-quantitative examinations (as well as hopefully reaping the fruits). Actually assessing \(L\) (as a quantified value) in any particular case may be next-to-impossible: it goes beyond the whole bewildering complexity of a creature’s entire behavioral space in its environment (which also includes other agents), as it also involves new deeper interactions between two such ones. Cognitive constraints and others determine how close the union could come to a truly single agent (overcoming coordination problems, for example). Even so, simply forming elementary mathematical equations with \(L\) as a parameter, and examining relations between their parts, may be a valuable normative analysis of the principles underlying the phenomenon and governing its many strategic faces.\textsuperscript{69} It may serve as a confirmation for the reasonableness of this theory, as well as point to the ’right’ questions that should be asked about the phenomenon, and provide further insights into its perplexing phenomenology.

For starters, the parental investment required is obviously a central parameter on which \(L\) depends. Not just on its general level, but also its time-length and distribution, the type of support, and other factors. The human species is an extreme outlier here\textsuperscript{70}, in the general level of support required, and particularly in its duration, which can \textit{far exceed the time between childbirths}\textsuperscript{71}. This gives an advantage to sticking with the same partner over multiple births, which allows for a fuller unity with a more substantial synergetic value (e.g. no conflict regarding who to bring food to, including competitive manipulations on the part of the different co-parents). In contrast, for species where intense but short bursts of investment are required, it may be possible for partial unity to arise and then later dissipate naturally: Given very high death rates (say), that make it \textit{unlikely} for a parent to live long enough to reach a second spawning season, the situation is somewhat similar to that of the single-copulation species; the parents are heavily-invested together in their central project. However, \textit{given} that they do manage not only to outlive the offspring’s dependence period but to reach another season, they may again become free to pursue their own individualistic interests. Their contextualized utilities now fall apart, while this future wasn’t a priori likely enough to drive a serious wedge between the united before.

Other such connections and questions abound:

- How does the overall value of falling in love change along the spectrum whose ends we’ve depicted (single-copulation vs. Don Juan)?
- Or the value of being in love, along the spectrum of having the whole reproductive future ahead vs. having no prospects for more children but only the existing ones to take care of?

\textsuperscript{68} “And what would humans be without love? 
RARE, said Death.” –Pratchett

\textsuperscript{69} Colyvan (2010) focuses on love from this strategic point of view, furthermore touching upon the discrepancy between humans’ familiar ’moralist’ view of love-related issues, and this ”cynical”, economics approach to analyzing such behavior and feeling.

\textsuperscript{70} While admitting significant cross cultural and environmental variation.

\textsuperscript{71} Thanks to Eyal Winter for pointing this out.
When copulation doesn’t usually result in impregnation, but multiple tries over a lengthy period of time are required, how does this effect the development of love (instantaneous vs. gradual growth)?

Falling in love one-sidedly, with someone who doesn’t seem to be undergoing the matching process – is it mostly just a misunderstanding or some other miscalculation? Or could it be a risky strategy, tempting the other side by making it easier for them to glide into the union with much less risk?

Given that falling in love can also be faked, what would the strategies regarding this choice look like? How would they change as the other parameters change? Specifically, given that there’s some (parameterized) chance of the fraud being detected? (Detection thereby influencing the other side’s preferred strategy.)

How should the phenomenon alter relative to the cultural and historical differences in romantic conducts? For example, does love (within the family) really vanish in societies where the match is not only approved but actually made by the happy couple’s parents? (The theory would suggest, quite controversially perhaps, that this would be far from the case.)

And many, many more.

### 6.5 Beyond Heterosexual and Romantic Love?

**Section Summary**

Related types of love and other symbiotic relationships shed some more light on love as well as on the framework.

Alternatively to our working assumptions and account, one could forget about psychology being shaped according to the genes’ interests (in this strong fashion of evolutionary psychology), while still try to ground love in the synergetic value of unity – now perhaps serving the organism itself. There is certainly something to that. Friends or business partners can work in harmony and become to some extent co-dependent as they become more efficient as a single operating unit for the benefit of them both (we may even have our innate social faculties recognizing and supporting such cooperations). What this lacks, what the genetic situation in sexual reproductions adds, is the drastic mechanism that forces unity through brute identification of interests. (This, of course, is basically why love comes in pairs). The best of friends is still far more replaceable than an existing offspring’s other parent. The variety of other life situations almost never gives us this extent of interest-identification; and certainly not with this level of evolution-scale regularity, which is what leads

---

72 "Displays of love provide another sign of commitment. A man can attract a woman by doing special things for her, showing a loving devotion to her, and saying "I love you." Men and women rate these tactics among the top 10 percent of all tactics for attracting a permanent female mate. Demonstrations of love convey cues to long-term commitment." (Buss, 2011) p. 104.

73 This is in a sense the path our nearest neighbor, Frank (1988), takes, working within a framework of ultimately individualistic (organism) rationality. One could formally add biases to the organism’s preference to get equivalence to the gene-centric view (this is probably what he (and Pinker) have in mind). But following Dawkins, we find it cleaner and indeed conceptually correct to just consider "the genes’ interests" and "the genes’ interests as approximated through their vehicle’s psychology". There is no independent sense to the individual’s interests – western-culture individualism notwithstanding.

74 On this reading, group-selection is generalized properly: groups small enough, similar enough in interest be better off in unity, do enter into unions and are selected as such. It is groups too large for this, where inner-group selfishness raises its head, to which the modern Dawkins (1976) critique of group selection applies.
to the architecture of mind to support it. So there certainly may be (and there do seem to be) innate mechanisms particular to parents-bonding, following the principles that our normative analysis dictates. But our framework, recall, is a general one, of the psychological level constructed to follow the genes’ interests. It is just that it responds appropriately to the evolutionarily-dramatic situation itself – far more strongly perhaps than our culturally-influenced conscious level would have it. (The conscious level might be misled as for the true interests, as for example in putting the individual organism at the center).75

**Blood Ties**

These are another sort of relationships involving a similarly deep, genetically-grounded fate-sharing. In principle, our general framework should apply here just as much. But the details do not seem to allow for the application to come out as striking: Siblings’ genomes, although sharing an average of 50% common fate (‘inclusive fitness’), have the rest of the percentage in an inherent state of considerable competition (for parental investment and other resources). And parent-offspring genomes, at precisely 50%, are in a deeply asymmetric relationship, in terms of their vehicles’ current stage in life (their level of independence, the prospects for future offspring, etc.). This, on the one hand, means far less competition (parents tend to collect food for their young ones rather than struggle against them over it) (though parent-offspring conflicts can certainly get nasty). But on the other hand, it breaks the symmetry of love itself. For example, parents may seem to take care of their offspring ‘for their own sake’, as the vehicles of their reproductive path (the offspring’s own personal success being what the parents also get in return); and because the offspring rely on that support critically. Meanwhile, the beneficiary offspring may care about their parents first and foremost for their help’s sake (their support of the other siblings coming a distant second). And the general implications for mental unity, for beings at such different places in life (and with brains at such different stages of development), are far less obvious (and less principled). We have thus kept our focus exclusively on romantic love. But what of cases where romantic love seems uncalled for?

**Old Age**

Our logic suggests that old people, past the age of fertility, should not fall in love, but rather focus on taking after their existing children & grandchildren instead. To some statistical extent this is certainly the case, but not completely. However, the general sad rule for the exceptionally-old (by nature’s standards) applies here too: It is too rare or inconsequential. Evolution doesn’t care. Just like it can let bodies fall apart, it can keep intact mechanisms that were fit for fertile maturity. Old age does not drive adaptation.

---

75 The following related analogy to physical pain & pleasure may perhaps be helpful: We have a rich predictive-reactive psychological system for getting us what we want – which is what evolution wants. Its boundary of physiological pain & pleasure determines many basic needs and wants: nutritious food, bodily unharm, sex, etc. The levels of the various pains & pleasures have to be balanced proportionally – so that we would prioritize correctly. So, likewise for prioritizing co-parenting over life’s other interest-sharing situations (which are more partial and much weaker). And just as there is also room for shallower innate instinctive mechanisms (than conscious bodily harm avoidance), some further innate mechanisms may exist particularly for the central and prevalent interest-sharing involved in mating.


77 In some languages, these different types of love are indeed designated by different words.

78 I am not aware of relevant empirical data, and gaining insight through such statistics would be tricky anyhow.
Homosexuality

Homosexual love should not be explained away as swiftly. At least on the face of it, it is not simply (evolutionarily) irrelevant, but actually anti-adaptive. (For example, harming by directing caring attention to the lover, away from an existing spouse of opposing sex and the co-created kids). But love or no love, homosexuality itself as a sexual preference (as mentioned in section 5.3) is (seemingly?) as anti-adaptive as it gets (regardless of parental investment). Obviously we will not attempt to offer anything for this general outstanding riddle in evolutionary theory. And we cannot account for homosexual love before homosexuality is adequately explained (for example by showing how in fact it can be adaptive). We would hope for fully incorporating our approach to love with that explanation of homosexuality when it does arrive. For now, there is an intriguing connection between the two that we can, very speculatively, already point to:

Apparently, then, something (innate / genetic?) can "go wrong", misdirecting the sexual preference. And the potential to fall in love, too, may go astray, turned towards the wrong sex. Our question is: Why are the two so perfectly correlated in their divergence? Why at all? On the face of it, there's no conceptual necessity here. One could conceivably be in love with his wife (exhibiting and experiencing its many non-sexual aspects) while being sexually attracted to only men. Here, our theory's principled directed connection, from sex to love, may suggest the following explanatory picture:

Sexual preference (at its most general) is supposed to capture the potential for reproduction, in the broadest biological terms (ruling out the wrong sex, as well as other species etc.). There is some system for recognizing that potential (a sub-system within the larger one determining sexual attraction in all its detail – including responses dependent on the particular other individual and the situation). That sexual preference system is evolutionarily extremely primal, there for almost all creatures – even those too simple to cognitively handle particular identities of fellow ones. How does love figure into this picture?

On the way to proliferation, sex is a central necessary step. Sexual preference determines the potential for shared offspring (rape excluded), and through it the potential for (being better-off) sharing a soul (as our analysis would dictate). So there's not much point in falling in love with someone who isn't even a potential for wanting to have sex with. The attachment mechanism can thus latch on to the sexual preference system (which is already there anyway) for the purpose of recognizing who is a sex-potential. Depending on the cognitive structure, on how the modules are put together, this hook-up may be possible from the outset (responding to the biological type), before involving the intricate personal-level (strategic and cognitive) complexities between sexual attraction and love.

If this is indeed the setup, if people are aimed at falling in love only with those they can be attracted to biologically, then that perfect correlation in homosexuality can be easily explained: Whatever it is that's going wrong, happens within the sexual preference system itself; and the attachment mechanism, intact, but not in charge of that aspect, just follows.

---

79 As usual, the suggested normativity is evolutionary, naturalistic – not at all moralistic.
6.6  Interim Conclusion

This chapter finally connected the abstract framework and strategic principles to the actual phenomenon in our lives, demonstrating how our general approach frames the particular exploration\textsuperscript{80}. Many details may surely require revisions, taking cues from the improving scientific understanding of the actual inner-workings etc. We believe the value of this work is first & foremost in the theoretical framing itself.

We've come a long way to ultimately end up with a bottom line Aristotle already formulated more than two thousand years ago. Our approach to love has put agent-unification at the very heart of love, constituting the phenomenon. But is love in all its glorious complexity really exhausted by this? Does our framework truly capture the essence of love, through which it could fully be accounted for? Making the complete case would require figuring in each and every aspect of love, each topic that has something substantial to do with it. This would be the overarching purpose of Part II.

\textsuperscript{80} as I haven’t seen in the literature, although the strategic analysis of behavior & cognition certainly stands at the core of evolutionary psychology.
Bibliography

Aumann Robert Who are the players?. - 2012. - Lecture at the 20th-celebration of the Center for the Study of Rationality.


Schopenhauer Arthur Metaphysics of Love [Report].


