Is High Fidelity Imitation Necessary for Cumulative Culture?

An Analysis of the Ratchet Argument

Chiara Elettra Ferrario

A thesis submitted for the degree of Doctor of Philosophy Victoria University of Wellington January 2018

Under the supervision of Professors Richard Joyce (Victoria University of Wellington) Kim Sterelny (Australian National University)

II

Statement of Originality

This thesis is the sole work of its author; no one of its part has previously been submitted for any degree, or is currently being submitted for any other degree. To the best of my knowledge, this thesis contains no material previously published or written by another person, except where due reference is made in the text.

Chiara Elettra Ferrario

Acknowledgments



VI

Abstract

The so-called *ratchet hypothesis*, advanced in the 1990s by developmental and comparative psychologist Michael Tomasello, is the source of one of the most commonly reiterated platitudes surrounding evolutionary explanations of culture. The gradual, ever-increasing (and never *decreasing*) build-up of knowledge, skills, objects and innovations that accompanies our everyday existence is an unprecedented and unmatched human peculiarity—something that sets us apart from the rest of the biological world. This extraordinary form of culture, which is defined as *cumulative*, is attained by humans in virtue of our unique capacity to *faithfully imitate* each other.

In this dissertation, I revisit the origins and development of this peculiar and extremely influential argument. I show how the original hypothesis by Tomasello has been progressively simplified and standardized to become an intuitive and extremely prolific (yet ironically not quite accurate) form of orthodoxy in cultural evolution studies and beyond.

I then proceed to deconstruct this line of reasoning into its fundamental components: namely, *imitation*, *high fidelity* and *cumulative culture*. The conceptual and historical analysis of each term and of their theoretical connections will show that the simplicity and uniformity of the ratchet argument is only illusory. In reality, this argument allows for a bewildering variety of interpretations—resulting, in fact, in a multiplicity of slightly divergent "ratchet arguments". My investigation will help to clarify why the argument elicits starkly opposed reactions, from enthusiastic endorsements to categorical dismissals. It may also assist in dispelling several (apparent) long term controversies. As for the reappraisal of the argument's ultimate significance to evolutionary explanations of cumulative culture, I envisage a very narrow applicability. Of the many ratchet argument interpretations isolated in this work, only a minor proportion have something meaningful to say about the evolution of complex, cumulative cultural phenomena.

Table of Contents

Acknowledgments	V
Abstract	VII
Table of Contents	IX
List of Figures	XIII
List of Tables	XV
CHAPTER 1. The Ratchet Argument as Orthodoxy	3
1.1. Introduction	3
1.2. Standard outline of the Ratchet Argument	6
1.3. Making a case: the Ratchet Argument as orthodoxy	11
1.4. The Ratchet logical apparatus: dropping attribution claims	20
1.5. Theoretical links in the Ratchet Argument	24
1.5.1. Why the links have been interpreted as necessity relations: enl	hancing
empirical tractability	24
1.5.2. Theoretical links as necessity relations: the drawback of explan	atory
gaps	30
1.6 Accessory claims	34
1.7. Conclusion: not one, but many ratchet arguments	39
CHAPTER 2. What is Imitation?	47
2.1. Introduction	47
2.2. Defining imitation and its place in social learning	48
2.3. Imitation: brief history of an idea	53
2.4. Historical ambiguity: all copying vs. a specific way of copying	60
2.4.1. Imitation vs. local enhancement	60
2.4.2. Imitation vs. emulation	63

2.4.3. Input or output?	66
2.5. From copying to learning: Perspectives from psychology and cult	ural
evolution	68
2.6. Interdisciplinary ambiguity: the channel vs. mechanism distinctio	n 73
2.6.1. Confusing channels and mechanisms	78
2.6.2. The relationship between imitation-mechanism and imitation	า-channel
in the Ratchet Argument	
2.7. Intra-disciplinary ambiguity about imitation	
2.7.1. True imitation, or mentalised imitation	90
2.7.2. Simple imitation, or mimicry	92
2.7.3. Over-imitation	95
2.7.4. Program-level imitation	97
2.8. Conclusion	100
CHAPTER 3. What is Cumulative Culture?	103
3.1. Introduction	103
3.2. Plain culture: one evolutionary perspective, many referents	110
3.3. Historical notions of cumulative culture	117
3.4. Contemporary notions of cumulative culture	123
3.5. Two competing definitions of cumulative culture	125
3.5.1. Jablonka's account: accumulation of traits	126
3.5.2. Tomasello's account: accumulation of trait's modifications	129
3.6. Complexity	132
3.6.1. Three kinds of complexity	133
3.6.2. Understanding interdependencies	135
3.7. Efficiency	142
3.7.1. Missing efficiency?	145
3.7.2. A different kind of efficiency: efficiency and S-functions	151
3.7.3. Efficiency, fitness and viability in technological traits	157
3.8. Recapitulation and cognitive complexity	159
3.9. Conclusion	163
CHAPTER 4. What is Fidelity?	173
4.1. Introduction	173

4.2. Competing notions of fidelity
4.2.1. Sperber and action matching17
4.2.2. Laland and trait preservation18
4.3. Levels matter: levels discrepancy in fidelity and problems for the Ratchet
Argument19
4.3.1. Fidelity as similarity: action matching and trait preservation 19
4.3.2. The Ratchet Argument adopts inconsistent meanings of fidelity. What
to do about it?19
4.4. Reconciling L1 and L2: bridging the gaps between levels
4.4.1. Recent work on the Ratchet Argument neglects the levels discrepance
4.4.2. Action matching is not trait preservation and doesn't naturally
correlate with it
4.4.3. One proposal to bridge the gap between low-level and high-level
fidelity; a third account of fidelity emerges
4.5. The argument from biology: fidelity as inheritance
4.6. Models of inheritance: replication vs. reproduction plus inheritance 21
4.6.1. The roles of inheritance and similarity in evolution
4.7. Weismannian inheritance and the gradualist interpretation of the Ratcher
Argument
4.8. Conclusion
CHAPTER 5. Conclusion
5.1. Introduction
5.1.1. Terminology recap and filling the Abacus diagram
5.2. The Gradualist Interpretation
5.2.1. Adding depth to imitation: a speculative tweak to the gradualist
interpretation, or the Exaptive Interpretation
5.3. The Social Norm Interpretation
5.4. Quantitative Interpretations
5.4.1. Other quantitative interpretations: a beam of redemption
5.5. Conclusions
References

List of Figures

Figure 1. A graphic outline of the minimal RA's apparatus
Figure 2. Three ingredients or terms in the minimal RA
Figure 3. Two explicit theoretical inks in the minimal RA
Figure 4. A third implicit theoretical link via transitivity
Figure 5. Imitation is necessary to high fidelity; high fidelity is necessary to
cumulative culture
Figure 6. Including attributions claims: postulating imitation as uniquely human
Figure 7. Illustrating explanatorily defective necessity relations: the fire triangle
Figure 8. The Abacus Diagram. Prefiguring competing interpretations of the RA.
Figure 9. From Whiten & Ham (1992. Here in abridged version from Galef (2012),
excluding non-behavioral mimetic processes59
Figure 10. Schematic conceptualization of a prototypical social learning event. 66
Figure 11. Different categorisation paths examined in this chapter and the
resulting accounts of imitation in the RA 101
Figure 12. Three banner cases of cumulative culture
Figure 13. Artefacts morphologies yield gradual function-based lineages linked
by ancestry, in Pitt-Rivers work collected by Balfour 119
Figure 14. The evolution of forks 142
Figure 15. Examples of "uncomfortable" (and inefficient) design: the thick fork,
the masochistic coffeepot, and the slipper spoon
Figure 16. From genotype to phenotype (adapted from Bergstrom & Rosvall
2008)

Figure 17. Phenotype to phenotype inheritance may be less faithful, but is	
sufficient for evolution by natural selection 2	230
Figure 18. Some possible interpretations of the RA on the Abacus Diagram 2	241

List of Tables

Table 1: RA compendia in the literature	12
Table 2: Evolutionary definitions of "culture"	111
Table 3: Fidelity is conceived at three different levels	234

CHAPTER 1. The Ratchet Argument as Orthodoxy

1.1. Introduction

In 1993 Michael Tomasello and colleagues put forward an argument that would profoundly affect the field of social learning and cultural evolution studies for more than two decades (Tomasello et al. , 1993). The argument concerned the role of a specific social learning mechanism, imitation, in the evolution of cumulative culture. It became popular under the label the "ratchet effect" or "imitation hypothesis" (Morin, 2015). In the ratchet effect hypothesis, *cumulative culture* is regarded as a human distinctive trait. Culture is a product of social learning, but cumulative culture can evolve only when the social learning mechanism transmits behaviour and other cultural entities (such as products of behaviour, i.e. material or linguistic artefacts) with sufficient *high fidelity*. This high fidelity mechanism is identified with imitation. Together, these assumptions have established a unique connection between cumulative culture, imitation and humanity.

In this chapter, I will offer an initial analysis of the ratchet hypothesis' structure, origins, and reception. Although the idea was originally put forward by Tomasello (Tennie, Call, & Tomasello, 2009; Tomasello et al. 1993) and gained popularity under his name, it has been reported, promoted or discussed in a number of similar or directly derived forms. I will refer to these spin-offs of Tomasello's original idea as "ratchet arguments". Despite the recognised single source, I argue that what is habitually understood as a "ratchet argument" in the literature differs in subtle but important ways from the original proposal by its author. One glaring difference, for example, rests in the prominence accorded to imitation. The seminal

1993 paper attributed a less central role to imitation, proposing "cultural learning" as the mechanism responsible for the unique evolution of complex culture in humans. Imitation (or "imitative learning") was in turn one of three essential components of cultural learning, along with "instructed learning" and "collaborative learning", offering thus a more fine-grained picture of the abilities required for cumulative culture. This picture also stressed the role of "intersubjectivity or perspective taking" (Tomasello et al., 1993: 495), prefiguring the current theoretical orientation of Tomasello's research, which has virtually abandoned the "imitation hypothesis" in favour of the importance of "shared intentionality" mechanisms (Tomasello, 2014; 2009). Somehow, however, this nuanced picture failed to make its way into the literature, and Tomasello's ratchet hypothesis became problematically entwined first and foremost with imitation.

Thus the "ratchet argument", as I intend it here, is a rather radicalised and in many respects simplified version of Tomasello's ideas; nevertheless, it displays a remarkable persistence and consistency throughout the literature, and is therefore worthy of consideration on its own. My aim in this first chapter is to identify clearly the boundaries and precise formulation of this target argument and to delineate its logical structure, which will serve as the blueprint for analysis in the rest of the dissertation. Throughout the thesis, I will refer to the target formulation that I am going to sketch in this section simply as the "ratchet argument" ("RA" from now on). This labelling is also intended to underscore the difference of this popularised narrative from the original formulation of Tomasello's ratchet effect.

Aside from its pervasive impact on the literature, there are several reasons to be interested in the RA. Here are the most obvious two. First, the RA has a highstake explanandum: cumulative culture. Cumulative culture is not only striking *per se*, but invites special consideration because it has the profile of a virtually exclusive human trait (cf. for a comprehensive overview Dean, Vale, Laland, Flynn, & Kendal, 2014). If it is true that spaceships, Google Scholar, smartphones and genetic engineering are possible because we are the only species that through imitation can make small, continuous, and cumulative improvements in our knowledge and skillset, then looking at how and why this process got started is a pretty interesting business. The question of what underpins the difference between human and animal cultures was listed by *Science* among the most important 125 unanswered questions of substantial import to how we view our place in nature (Anon, 2005; Dean et al., 2014).

Second, the RA builds on a central tenet of evolutionary biology: namely, that complex, cumulative adaptations and evolutionary products depend necessarily on high fidelity inheritance. The intellectual ancestor of the RA is, so to speak, Richard Dawkins's seminal argument about high fidelity replicators as a necessary condition for the evolution of life's spectacular complexity (Dawkins, 1976, 1986), which seeped into cultural reasoning via substrate-neutral evolutionary thinking (that is, via abstract descriptions of evolution by natural selection as an algorithm applicable in principle to any non-biological system displaying variation, inheritance and differential reproductive outputs; see Dennett, 1995). The connection between fidelity and cumulative evolution is a central topic to other evolutionary accounts, not necessarily connected with Dawkins' so-called "replicators framework"-evodevo and evolvability research programs, for example (Griffiths & Gray, 2001; Jablonka & Lamb, 2006; Laland et al., 2015; Sterelny, 2001b, 2004, 2006, 2011). Despite its centrality, however, very little work has been done explicitly on fidelity, and the notion appears in many respects ambiguous, especially in the context of cultural evolution. I will argue that we need to think more about the meaning of "fidelity" and its effects on cumulativeness when we export the notion to the cultural domain: several lines of argument suggest that the application of a "Darwinian algorithm" (Dennett, 1995) outside its native context of biology requires extensive reconceptualisation (Godfrey-Smith, 2009).

In sum, this first chapter shall show specifically that: (a) the RA represents a long-held form of orthodoxy; (b) this form of orthodoxy is still largely influential, especially among the numerous disciplines that surround specialised debates in social learning and cultural evolution, and therefore affects peripheral (and sometimes not so peripheral) theorisation; (c) this "ratchet orthodoxy" comes in the

5

form of a somewhat trivialised and simplified package of ideas with respect to Tomasello and colleagues' original hypothesis, but it has an identity and is transferred consistently through the literature; and (d) the package more often than not includes the elements of imitation, fidelity, and cumulative culture, connected in the argument by theoretical links that allow for different interpretations.

Ultimately, I will suggest that a reappraisal of the RA is worthwhile, both hermeneutically, to appreciate its historical impact on the literature, and theoretically, to evaluate whether it can still provide insights to the current debate on cultural evolution.

1.2. Standard outline of the Ratchet Argument

The set of Tomasello's papers habitually linked to and quoted as a source of the RA comprises Tomasello et al., 1993; Tomasello, 1990, 1999. Ratchet-like ideas gained resonance through a subsequent paper by Boyd and Richerson (1996) which emphasised the role of imitation (as opposed to other social learning modes), its adaptiveness, and the relevance of non-verbal transmission ensured by imitation, which makes it a valuable *explanans* in a comparative context, avoiding anthropocentric biases. Boyd and Richerson contributed to the popularity of the ratchet argument as a standard argument for the evolution of human cumulative culture as opposed to simple animal culture. In addition, their formulation emphasised the role of imitation (dropping Tomasello and colleagues' reference to cultural learning, language, teaching and shared intentionality), and anchoring the RA firmly to the mainstream debate on cultural evolution.¹ Most quotes of ratchet ideas indicate one or more of the aforementioned works as the sources of the argument.

¹ Boyd and Richerson's evolutionary theory initially characterised imitation as a cheap and easy form of learning (Boyd & Richerson 1985). In later work, however, Boyd and Richerson came to substantially endorse Tomasello's views of imitation, presenting it as a high fidelity transmission

As I noted, however, the package of ideas habitually identified with the RA over the last two decades has evolved to a form that is rather distinct from the original formulations. A canonical, minimal outline of standard RAs would include the following propositions:

- I. The human species shows, to a massive degree, a unique trait in the animal kingdom: complex or cumulative culture.
- Culture depends on social learning. But complex cumulative culture depends on some high-fidelity form of social learning.
- III. This form of social learning is typically identified with imitation, in virtue of its being the most (or only) high-fidelity type of extra-verbal social learning.

I think that these three statements capture the essential legacy of the RA. There are other aspects to it, but the three propositions summarise the take-home message that any reader would recognise as "the RA", with virtually no controversy. I will group secondary, derivative and more controversial aspects of the argument in a set of "accessory statements", below. As far as the minimal outline goes, notice that it encompasses *two* main theoretical nodes or links. Proposition (II) contains the first theoretical node: it postulates a connection between the properties of fidelity and cumulativeness or complexity in culture (the high fidelity / cumulative culture link). Proposition (III) states the second theoretical node: it associates imitation to high fidelity (the imitation / high

channel, and emphasising its potential to generate fine-grained and cumulative cultural adaptations (Boyd & Richerson, 1996; Richerson & Boyd, 2005). Their endorsement magnified the impact of these ideas. Despite imitation's playing a much less central role in their multifaceted picture, their theory is one of the most serious and respected attempts to apply evolutionary ideas to culture, and therefore has had a lasting impact on the ideas presented in this context too.

fidelity link). This association comes in a weaker form (imitation is the *most* high fidelity mechanism) and a stronger form (imitation is the *only* high fidelity mechanism). I have therefore decoupled the stronger version into an accessory claim below, stating that fidelity is unique to imitation—see (V) below. Proposition (I) motivates (so to speak) the argument: cumulative culture is a phenomenon that needs explanation, being exclusively represented in our species. I will suggest that a fruitful approach to the analysis of the minimal RA is to reduce its content to three fundamental notions and two (or three) main theoretical relations among them.

The set of accessory statements that are often associated with RAs (although to varying degrees and with differing relevance) can be outlined as follows:

- IV. Imitation is a uniquely human trait.
- V. Fidelity is unique to imitation.
- VI. Imitation is an adaptation for cumulative culture.
- VII. Inherent features of the social learning mechanism (fidelity in particular) are the primary factors responsible for the nature of the cumulative cultural phenomenon.
- VIII. The overall argument is often restricted or thought to properly apply to a particular subset of cultural phenomena, i.e. complex material culture, technology, and skill.

I will go through the accessory statements in detail in the section 1.6. Here, I want to call preliminary attention to a few interesting features of the bullet outline, in its minimal and extended versions. The first thing to notice is that

there are a few "claims of uniqueness" contained in the extensive version of the RA: cumulative culture is uniquely human (I); imitation is uniquely human (IV); and imitation is uniquely high fidelity (III, V). I will argue later that such "uniqueness" claims have standardly been interpreted as necessary conditions, although this doesn't always represent a fair construal, nor has it been done in a rigorous way (see sections 1.5.1 and 1.5.2). But consider for the moment only the claims of alleged human uniqueness. I will call these "attribution claims", because they attribute the trait in question exclusively to a particular species. When these claims are emphasised, the RA can be seen as joining a long list of arguments for the so-called "anthropological difference" (Glock, 2012; Oakley, 1957; Tomasello, 2014; Urban, 2014). These arguments aim at demonstrating that, evolutionarily speaking, some particular feature or property of humans sets them apart from other species. Such arguments have been used to provide support for a number of hypothetical watersheds between human and nonhuman animals (e.g. theory of mind, language, culture, symbolic thought and tool-making—just to mention a few). I suspect that the propensity to serve the anthropological difference cause contributes to the explanation of the extraordinary interest and diffusion of the RA. At the same time, it has diverted attention from the basic structure of the argument.

In fact, what is striking is that, in principle, the main theoretical nodes of the argument (i.e. the links established between imitation and high fidelity, and between high fidelity and cumulative culture by propositions III and II) could and should, I will argue—be assessed separately from the attribution claims. In other words, establishing whether imitation is an exclusively or exceptionally high fidelity mechanism (III), has nothing to do with whether imitation is exclusively human or not: fidelity of transmission should be explained from inbuilt features of imitation, not of humanity. Similarly, the content of claim (II), asserting that complex cumulative culture depends on some high-fidelity form of social learning, should be viewed as independent from the attribution of both cumulative culture and imitation to particular species. While the logical independency of attribution and theoretical claims in the RA may appear obvious, it is important to stress the point for it has been common practice to draw inferences from attribution claims associated with the RA, and especially from the alleged human uniqueness of imitation (see section 1.4).² The salience accorded to this or that trait's human exclusivity may well be the product of an anthropocentric bias that makes it intrinsically interesting to us; and I further suggest that the tendency to focus on (allegedly) human traits to do work in the RA may also derive from their relative empirical amenability. I discuss these dubious assessment strategies of the RA logic and their implications in greater detail in section 1.5.

The above "bullet outline" of the RA embodies an inevitable trade-off between simplicity of exposition and hermeneutical meticulousness. Despite my effort to rank claims into *central* and *accessory*, the characterisation will appear not enough or too inclusive to some, or failing to do thorough justice to the original formulation of the idea by Tomasello. But as I pointed out, the subject of this work is not the particular instance of Tomasello's detailed and (reasonably) well-identifiable 1993 argument, but rather a diffused, abstract and not necessarily accurate Tomasello-*inspired* orthodoxy, which flooded cultural evolution studies in the years following his seminal paper. Of this minimal and standardised orthodoxy, the bullet outline may be said to offer a fair representation. I now turn to document how these standardised reports are encountered in the literature.

² The minimal version of the RA is not strictly committed to the human uniqueness of imitation; it appears to suggest it under some interpretations, but it doesn't precisely imply it. In addition, the claim, which was originally very popular, has progressively become harder to maintain in the face of accumulating evidence concerning nonhuman species (Akins & Zentall, 1996; Bates & Byrne, 2010; Hurley & Chater, 2005; Kis et al., 2014; Nehaniv & Dautenhahn, 2007; Voelkl & Huber, 2000), so it would be a misrepresentation to say that most subscribers of the RA also endorse the view that imitation is a uniquely human trait.

1.3. Making a case: the Ratchet Argument as orthodoxy

What has come to represent the "standard version" of the ratchet argument is noteworthy because it has deeply influenced research trends in the field of cultural and human evolution for at least a decade, especially in the first years of its appearance (cf. for example Heyes, 1996; 1994a, 1994b; Sterelny, 2003; Zentall, 1996). The impressive amount of work on social learning and cultural evolution emerging in recent years has been considerably affected, more or less directly, by some version of ratchet ideas. Reactions have followed in a mixture of plain acceptance, adjustments, and more or less of severe critique, which I briefly review at the end of this section.

The main purpose of this section is, however, to show how the RA really is a consistent trope in the literature, and one that is often regarded as the received view about the evolution of cumulative culture. The most persuasive way to illustrate the reception and recounting of the RA through the years is to present a selection of quotes from different papers, ranging from the RA's first appearance to current years. Some of the quotes belong to RA's supporters, other to sceptics voicing the argument for critical purposes. As will be apparent, however, both advocates and detractors habitually converge on the same, simplified compendium of Tomasello's ideas that I have called "RA". The table below presents a chronological and fairly compact summary of RA "mirrors" in the literature, and is meant to highlight how the essential ingredients of the standard ratchet story (imitation, fidelity, cumulative culture, and more often than not human uniqueness) have evolved to become a solid package of ideas.

Table 1: RA compendia in the literature

Source	Quote
(Heyes, 1993)	"Some investigations [] (Galef 1992, Tomasello et al. in
	press) assume that cultures, or the products of cultures , are
	subset of traditions in which the focal behaviour, B, has
	been modified through 'the accumulation of modifications
	over time' (Tomasello et al. in press). According to this
	definition of culture, the potato-washing behaviour of
	Japanese macaques, [] would be counted as traditional
	but not cultural [] because there is no reason to believe
	that the present form of the potato-washing behaviour is
	the product of high-fidelity social transmission" (p.1004).
	"Thus, if information acquired through imitation is as likely
	as information acquired through individual or social
	learning to be lost or revised [], then imitation is no more
	likely than social learning to support culture" (p. 1005).
(Galef, 1998)	"Researchers working in different traditions seek evidence
	of imitation learning in animals for different reasons. []
	Furthermore, the cumulative culture that supports the
	biological success of humankind requires an ability to learn
	complex novel behaviours as a result of observing others
	exhibit them (Boyd & Richerson, 1996; Tomasello et al.
	1993). Consequently, evidence of an ability to copy
	faithfully complex, novel behaviour is important to those
	interested in animal traditions as homologous to human
	culture (Galef, 1992)" (p. 289-291).
(Mithen, 1999)	"Another recent suggestion has been that imitation is the
	critical mental capacity that allows human cultural

Source	Quote
	innovations to build upon each other and, thus, to produce
	cumulative cultural change (Boyd & Richerson 1996)" (p.
	408).
(Sterelny, 2003)	"Tomasello (1999, 2000) has argued that the evolution of a
	quite specific learning capacity plays a critical role in the
	transition to modern human behaviour: namely the
	evolution of true imitation [] True imitation makes
	cumulative improvement in technique possible."
	"[T]he paleoanthropological signal of high-fidelity true
	imitation is the cumulative multi-generational construction
	of new skills." (From Chapter 4).
(Avital & Jablonka,	"If, as has been claimed, only true imitation has a high
2000)	enough fidelity to allow cumulative cultural evolution, then
	it is only in our own species, where imitation is highly
	developed, that complex cultural adaptations can evolve"
	(p. 94).
(Richerson & Boyd,	"[O]nly imitation gives rise to the cumulative cultural
2005)	evolution of complex behaviours and artefacts. []. To the
	extent that observers can rapidly and accurately use the
	behaviour of models as a starting point, imitation leads to
	the cumulative evolution of behaviours that no single
	individual could invent on its own" (p.108-109).
	"Children imitate very faithfully, while apes emulate or at
	least imitate less faithfully" (p. 110).
(Caldwell & Millen,	"Many have drawn attention to the prevalence of
2009)	cumulative cultural evolution in humans, and noted that
	this particular feature appears to be either absent (Galef,

Source	Quote
	1992; Tomasello, 1999), or minimal at best (Boesch &
	Tomasello, 1998; Boyd & Richerson, 1996; Heyes, 1993;
	Laland & Hoppitt, 2003; Whiten et al., 2003) amongst
	nonhuman animals. The apparent rarity of this kind of
	learning amongst species other than humans has prompted
	much speculation about the possible reasons for its
	scarcity, the majority of which have concerned the probable
	learning mechanisms involved and the possibility that these
	may be unique to humans. Boyd and Richerson (1996) have
	argued that imitation ("learning to do an act by seeing it
	done", Whiten & Ham, 1992) may be necessary for
	cumulative cultural evolution. Tomasello (Tomasello, 1999;
	Tomasello et al., 1993) has proposed that both imitation
	and teaching provide the foundations of cumulative cultural
	evolution. [] Such interpretations are consistent with
	findings indicating that human children are far more precise
	imitators than are chimpanzees (e.g. Horner & Whiten,
	2005; Nagell, Olguin & Tomasello, 1993)" (p.1-2).
(Claidière & Sperber,	"Imitation in particular has been claimed to exhibit a
2009)	higher degree of fidelity and therefore play a major role in
	the evolution of culture " (p. 350).
(Whiten, McGuigan,	"Reviews of several later studies have tended to confirm
Marshall-Pescini, &	this picture, with chimpanzees failing to copy details of a
Hopper, 2009)	model's actions (especially in comparison to children's
	imitation), and performing instead in ways consistent with
	emulation (Tomasello & Call 1997; Whiten et al. 2004; Call
	et al. 2005)[]. Such low-fidelity copying was suggested to
	severely constrain the capacity of chimpanzees for the

Source	Quote
	cultural transmission of traditions (Tomasello 1999)" (p.
	2419).
	"The hypothesis that it is the lack of copying ability that
	prevents chimpanzees from generating the cumulative
	cultures so evident in the human case (Boyd & Richerson
	1996; Tomasello 1999) now sits less comfortably with these
	results" (p. 2425).
(Thornton & Clutton-	"One common argument for the lack of cumulative culture
Brock, 2011)	in non-humans is that human culture is underpinned by
	higher fidelity mechanisms of social learning than those
	prevalent in other species. One such mechanism is
	imitation, which was commonly thought to be rare or
	absent in non-human animals" (p. 984).
(Acerbi, Jacquet, &	"In particular, since imitation—i.e. the high-fidelity copy of
Tennie, 2012)	novel behaviours through the reproduction of action
	sequences of observed individuals (Call and Carpenter,
	2002; Tennie et al., 2006; Whiten et al., 2009)—has often
	been viewed as the learning mechanism that best explained
	the emergence of human traditions (Boyd and Richerson,
	1996; Tomasello et al., 1993; Tomasello, 1999), it is
	assumed that also non-human traditions are supported by
	similar imitative capacities (Claidière and Sperber, 2010)"
	(p.307).

Source	Quote
(Galef, 2012)	"The vast majority of analysed instances of social learning in animals appear to be the result of either local enhancement or emulation. [] There is no precise copying of the behaviour of proficient individuals and, therefore, no possibility of cumulative improvement in performance. Only precise copying of the behaviour of a succession of proficient demonstrators repeated over 'generations' can support the cumulative improvement that results from repeated grafting of adaptive innovations onto socially transmitted behaviours. Precise imitation, teaching, language, etc. each has the potential to promote precise copying and the consequent emergence of cumulative
	culture" (p.589).
(Heyes, 2013b)	"[I]t seems that imitation is more likely than emulation to result in high-fidelity transmission of behaviour- transmission of the kind needed for cumulative cultural evolution" (p. 323).
(Yamamoto, Humle, & Tanaka, 2013)	"However, many argue that humans are still unique in their capacity for cumulative cultural evolution, with successive generations building on earlier achievements [4–7]. This process depends upon faithful , high fidelity social transmission of improved, more efficient techniques. In humans, imitation and teaching are viewed as the key processes underlying cumulative cultural evolution" (p. 1).
(Galef, 2013)	"One of the more unusual characteristics of human culture is its cumulative nature, i.e. the ability of succeeding cultural generations to learn socially and then improve

Source	Quote	
	upon innovations developed by their predecessors. As	
	Tomasello (1994), Boyd and Richerson (1996), Tennie et al.	
	(2009) and others (for review, see Caldwell and Millen,	
	2008) have proposed, cumulative culture requires an ability	
	to precisely copy the behaviour of others (See also Galef,	
	1992). The same authors have argued that precise copying	
	of observed behaviour requires either pedagogy (deliberate	
	tuition) or imitation " (p. 125).	
(Wasielewski, 2014)	"Imitation, which is defined as behavioural replication in an	
	observer following witnessing of that behaviour in a	
	demonstrator, is proposed to be especially important for	
	the accumulation of cultural information because it allows	
	high-fidelity information transfer between individuals"	
	(p.2).	

It should be evident from the table how senior and junior scholars alike, across the last two decades, recognise the RA as a solid, established conceptual framework where imitation is generally regarded as the *only* social learning mechanism with high fidelity, or the one with the *highest* fidelity, and as such responsible for the emergence of cumulative traits in culture.

Advocates

Among those who cite the ratchet effect as a view that has been and still is influential are established scholars like Boyd, Richerson and Laland, as well as junior ones like Enquist, Wasielewski and Acerbi (Acerbi, Jacquet, & Tennie, 2012; Boyd & Richerson, 1996; Enquist, Strimling, Eriksson, Laland, & Sjostrand,

2010; Lewis & Laland, 2012; Richerson & Boyd, 2005; Shipton, 2010; Wasielewski, 2014). Evolutionary scenarios based on the RA feature prominently in Sterelny (2003) who suggests that the evolutionary advent of imitation might be behind the shift in the archaeological record from the extreme conservativeness of Oldowan lithic technology, to the more complex features of Acheulean technology, and perhaps even more in the diversification of shapes and materials seen in the last 150k years of the archaeological record. Other archaeological interpretations linked to the RA include the work of Shipton and colleagues (Shipton, 2010; Shipton, Petraglia, & Paddayya, 2009), who see Acheulean or Mode II lithic technology as the signature of imitation—among other psychological dispositions. Amid recent examples, Enquist, Lewis and Laland explicitly see their models as a confirmation of the argument put forward by Tomasello (Enquist et al., 2010; Lewis & Laland, 2012, see chapter 4). Specifically, their work illustrates the centrality of fidelity requirements for social learning, remaining silent on the question of imitation. Wasielewski, drawing on a novel definition of complex technical tasks (Acerbi, Tennie, & Nunn, 2011), describes imitation as a necessary component of their evolution and endeavours to provide experimental support (Wasielewski, 2014). A particularly important example to quote is the work of Galef, which is habitually reported as defining cumulative culture as the exclusive product of imitation, one of the central theses of the RA (Caldwell & Millen, 2009; Galef, 1998; Whiten, McGuigan, Marshall-Pescini, & Hopper, 2009).

Detractors

Along with general acceptance, there also have been important critical reactions. Some of the critical responses to the RA have tackled what I have called its "accessory hypotheses" (IV-VIII)—for example, the hypothesis that imitation may be an adaptation in the human species (Heyes, 2013b)—but a sizable number have also targeted its central tenets. Many have criticised the

18

centrality accorded to imitation, claiming that there is no reason to believe that imitation is better equipped to deliver fidelity than other social learning mechanisms (Heyes, 1993), or that other social learning mechanisms can achieve sufficient fidelity (Avital & Jablonka, 2000; Sterelny, 2012). Others have emphasised that the excessive focus on social learning (claim VII) is inappropriate, arguing that additional factors have a role in assuring the necessary conditions to establish cumulative culture (including fidelity of transmission, see Henrich & Boyd, 2002; Henrich, Boyd, & Richerson, 2008). One notable example is the case of loss of complex technology in Tasmanian aboriginal groups described by Henrich, who shows how the demography and size of the social learners' pool are crucial factors to the maintenance of a group's technological suite (Henrich, 2004).

Counterarguments that have focused on challenging the sufficiency, necessity or mere convenience of imitation as the chief social learning mechanism in the production of culture are too numerous to be mentioned in a principled way here (though Eva Jablonka may be named as an early champion of sceptical perspective on imitation; see Avital & Jablonka, 2000). It will suffice to say that they have tackled the question from an array of perspectives: using empirical, experimental, anecdotal and ethnographical evidence. By and large, these counterarguments can be said to have re-sized the centrality accorded to imitation. The role of high fidelity social transmission has instead prompted less harsh doubt (Andersson, 2013; Caldwell & Millen, 2009; Heyes, 2013b), with the notable exception of the Sperberian and cognitive anthropology's school of thought, and the noteworthy case of Morin (Atran, 2001, 2002; Boyer, 1994, 1998; Morin, 2015; Sperber, 1996, see Chapter 4).

In sum, it may be fair to say that ratchet ideas have undergone substantial refinements and encountered an array of serious opponents. However, even if their importance to the field of cultural evolution has been curbed by competing theories, their legacy is alive and still influential, as it is evident from recent citations in Table 1. This may not be apparent to cultural evolution insiders, but

19

various versions of ratchet ideas are still the most popular account in outer circles of philosophy of biology, archaeology, psychology and other areas of the social sciences that take a tangential interest in the question of cultural evolution, though dealing less specifically on the cutting-edge debates (cf. Acerbi et al., 2012; Caldwell & Millen, 2008a, 2008b, 2009; Wasielewski, 2014).

My aim in this section has been to show that the RA: a) comes in the form of a package of ideas attributed to Tomasello, and Boyd and Richerson, but which eventually got simplified; b) consistently involves the three themes of imitation, fidelity, and accumulation; and c) represents a form of long-held orthodoxy whose legacy is at least in some interesting sense still alive. As such, it is worth of attention.

1.4. The Ratchet logical apparatus: dropping attribution claims

In the previous sections I suggested that the argument originally put forward by Tomasello and colleagues has been reduced over time, and (ironically) because of low-fidelity transmission, to a canonical, minimal form that is widely recognised as "the RA". I've also observed that the argument can include one or more "accessory statements". While I will briefly go through these accessory statements in section 1.6 and pick them up occasionally in subsequent chapters, most of this work will be concerned with the first three statements of the minimal outline. Thus in this section, I shall examine the logical articulation of these three fundamental claims, and explain how I intend to go about investigating them in the rest of the dissertation. The logical articulation of the three fundamental RA claims can be visualised as follows:



Figure 1. A graphic outline of the minimal RA's apparatus

The diagram illustrates concisely the same information contained in the verbal outline. Specifically, link 1 corresponds to proposition III, which establishes a connection between imitation and high fidelity; link 2 corresponds to proposition II, establishing a connection between high fidelity and cumulative culture. The graphic outline makes apparent how there are three main "ingredients" to the argument: imitation, high fidelity and cumulative culture.



Figure 2. Three ingredients or terms in the minimal RA

The three ingredients are connected by two explicit theoretical links: link 1 (imitation / high fidelity) and link 2 (high fidelity / cumulative culture). In the rest of this work, I will refer to these theoretical links respectively as L1 and L2 for brevity.



Figure 3. Two explicit theoretical inks in the minimal RA

I will argue in the following section (1.5) that these links are subject to different interpretations, but it is common to read them as necessary conditions. Hence
the information conveyed by the links is interpreted respectively as "imitation is necessary for high-fidelity" (crudely: "you don't get high-fidelity transmission without imitation"); and "high-fidelity is necessary for cumulative culture (likewise: "you don't get cumulative culture without high fidelity"). The necessity-based interpretation of the theoretical links brings about a third possibility. Necessity relations are transitive in nature, so from L1 and L2 a third link can be derived ("L3" from now on), establishing a direct necessity relation between imitation and cumulative culture:



Figure 4. A third implicit theoretical link via transitivity

This third derived link says, in a slogan, "no cumulative culture without imitation"; or "imitation is necessary to cumulative culture", which is in effect one of the most common interpretations of the RA message, as well its most obvious conclusion. It is worth considering the connection established between imitation and cumulative culture as a third link on its own, for the RA is frequently taken to establish a direct connection between imitation and cumulative culture, with fidelity playing a background or no explicit role. In this dissertation, the original L1 and L2 are given precedence, with the notion of fidelity being granted considerable space (chapter 4). However, I will show that the abridged connection of L3, bypassing fidelity, suggests an intriguing hypothesis in which imitation is either necessary for or particularly conducive to cumulative culture, but *not because of* its high fidelity—or at least not primarily because of its high fidelity (see section 5.2.1).

One may point out that the verbal outline of the RA starts with asserting the human uniqueness of cumulative culture, and wonder where that claim sits in the diagram. The answer to this worry is that the attribution claim about cumulative culture doesn't do any theoretical work in the argument. Likewise, the alleged human uniqueness of imitation is not represented in the diagram, for I maintain that attribution claims do not have any bearing on the logical structure of the RA. Hence, they are not included in the diagram, which simply lays out the relative positions of the terms and their logical connections. However, attribution claims directed at cumulative culture and even more at imitation did have an impact on interpreting the nature of theoretical links, and not a negligible one. I think that the unwarranted proclivity to incorporate such claims in the logic of the argument has biased the interpretation of the links towards necessity (see section 1.5.1); in fact, it is only under this interpretation that attribution claims can be considered active parts of the inferential machinery. However, the necessity interpretation is not necessarily the most convincing one, and in next section I will show that other interpretations are possible, most notably in terms of probabilistic relations established among the terms but also on a different epistemic and explanatory level.

I have chosen to represent imitation and cumulative culture in blue, with fidelity in orange, for the former terms represent more empirically investigable components of the argument. We may disagree on what we mean by imitation or cumulative culture, but we typically think of these two terms as actual traits in biological populations, something that we can go and observe—at least once the relevant definitions have been agreed upon. Fidelity is a different kind of object, an abstract property rather than a concrete entity. Therefore, since we have not only to agree on what "fidelity" means, but also on how it tangibly manifests, the problem of identifying what is meant by "fidelity" is in many ways more demanding than for the other two terms. In fact, I think that the elusive nature of fidelity has tended to distract readers from the proper logical articulation of the argument, which is fully captured by L1 and L2, and to encourage the unwarranted incorporation of attribution claims about imitation and cumulative culture (which are more empirically tractable) in the logical machinery. But I will say more about this soon.

The diagram in Figure 1 represents the basic layout that I am going to rely on in the rest of the dissertation, together with what I shall call the "abacus diagram" presented at the end of this chapter (section 1.7). I will now turn to discuss in greater detail the nature of theoretical relations in the standard RA.

1.5. Theoretical links in the Ratchet Argument

1.5.1. Why the links have been interpreted as necessity relations: enhancing empirical tractability

How shall we think of the two theoretical links in the standard RA? I will argue that they have been standardly and roughly interpreted as necessity relations. Other interpretations are possible, but the necessity one has been by far the simplest and most common. I will for a moment leave aside the matter of how correct the mainstream interpretation may be, and focus instead on explaining why I think this has been the default interpretation.

A simple way of conceiving of necessity relations is through a series of nesting sets. Each nesting subset depends on the existence of the encompassing set. The basic RA diagram in Figure 1 could therefore be re-represented as follows, to highlight necessity relations among terms, with the red and orange arrows representing L1 and L2 in the form of claims of necessity:



Figure 5. Imitation is necessary to high fidelity; high fidelity is necessary to cumulative culture

There are several cues suggesting that the standard interpretation of the theoretical links is in terms of necessity relations. (Even the derivation of L3 from L1 and L2 by transitivity may be seen as such a cue.) But particularly interesting and convincing evidence for this lies in the nature of the inferences drawn about the argument, and in the assessment methods that have been applied to it. These assessment methods have erroneously focused on the attribution claims regarding imitation and cumulative culture (i.e. whether these traits are exclusive to the human species or shared by others) and on exploiting counterexamples to these claims to draw inferences about the overall logic of the RA.

I have already remarked that a correct interpretation of the RA's logic should see attribution claims as independent from the internal logic of the argument. However, this is not what has happened historically. I argue that the only way in which empirical attribution claims about imitation or cumulative culture could be relevant to the RA is under the assumption that the theoretical relations represent a nested set of necessary conditions. Historically, attribution claims have been considered integral to the argument: this is evident from the impulse prompted by the RA for the empirical quest for imitation and cumulative culture in nonhuman animals. (Consider by way of example this non-exhaustive list of works concerned with determining nonhuman imitative and cultural capacities in the light of RA considerations: Akins & Zentall, 1996; Bugnyar & Huber, 1997; Byrne & Tomasello, 1995; Dean et al., 2014; Galef, 1998; Gergely & Csibra, 2006; Hunt & Gray, 2003; Jones, 2007; Kis, Huber, & Wilkinson, 2014; Pepperberg, 2005; Voelkl & Huber, 2000; Whiten, 1998; Whiten, Custance, Gomez, Teixidor, & Bard, 1996; Yamamoto et al., 2013; Zentall, 1996; Zentall, Sutton & Sherburne, 1996.) But such an impulse appears unmotivated *unless* the argument is interpreted as a set of nested necessary conditions, starting with postulating the human uniqueness of imitation. In sum, the necessity-based interpretation of the RA links is the most compelling justification for the empirical efforts to locate imitation and cumulative culture prompted by the RA.

Interpreting the RA links as necessity relations has interesting consequences, again related to the possibility of including attribution claims in the argumentative apparatus. To see more clearly how this works, consider that human uniqueness can be thought of as a necessary condition (only humans can imitate = humanity is necessary for imitation), and represent this as a further set encompassing the subsets depicting the argument's logic as per Figure 5. Postulating the human uniqueness of imitation requires adding a further circle, encircling imitation, to Figure 5:



Figure 6. Including attributions claims: postulating imitation as uniquely

human

As I said, claim (IV) about imitation has become increasingly difficult to maintain, and I thus place it among the accessory claims. However, the claim that imitation is a uniquely human social learning mechanism was strongly suggested by the minimal formulation of the RA, so it makes sense to think that it was readily incorporated in many simple formulations of the argument. If we hypothesise the theoretical links in the RA relations as a set of nested necessary conditions, *and* postulate in addition the human uniqueness condition *for imitation* as a further necessity relation, it suddenly makes sense to use empirical findings about imitation and cumulative culture's attribution to shed light on the other (allegedly) necessary claims in the argument.

In particular, the argument becomes suddenly susceptible to a certain class of counterexamples. For example, finding imitation in nonhuman species would seemingly open the way to the observation of cumulative culture in other species which had previously been considered to lack it: as noted before, the rush to find imitation in nonhuman animals has often been a product of such a perspective. Even more important, finding cumulative culture in nonhuman species that do not imitate would represent a counterexample to the RA, refuting the legitimacy of the theoretical links. (An example of this may found in Hunt & Gray, 2003.)

Again, the *correct* assessment strategy for the RA would call for leaving attribution claims out of the business and concentrating instead on establishing *how* and *why* L1 and L2 hold (or why they don't). Why, then, has the mistaken strategy of incorporating attribution claims in the form of necessary conditions been preferred? I think there are two connected reasons for this (beside the intellectual lure we have for finding—and rejecting—the ultimate divide between ours and other species—a temptation that makes attribution claims unusually irresistible).

First, the aforementioned strategy enhances the empirical tractability of the argument. Attribution or uniqueness claims in the form of necessity claims are relatively easily investigable, being subjected to empirical counterexamples

(i.e. presence or absence of imitative and cumulative cultural traits in nonhuman species, which is at least in principle testable, and has certainly been treated as such, cf. Bugnyar & Huber, 1997; Caldwell & Millen, 2008b; Hunt & Gray, 2003; Kis, Huber, & Wilkinson, 2014; Voelkl & Huber, 2000). A single empirical counterexample represents an accessible piece of information; yet it can invalidate the whole relation, thus providing maximum inferential return for minimum or limited epistemic effort. (Notice, however, that counterexamples can work only to invalidate, and not to confirm, the theoretical links). In sum, making the argument amenable to the counterexample strategy is attractive, for it immediately increases net returns of empirical investigation.

Second, having attribution claims do the inferential work in the argument via empirically collectible counterexamples saves one from dealing directly with the explanatory details of L1 and L2, which crucially involve fidelity. Imitation and cumulative culture (as I noted in section 1.4) represent more concrete items than fidelity of transmission: it is much harder to devise the correct empirical setting and the appropriate empirical counterexamples for claims involving the latter than the former. What represents a case of low or high fidelity transmission, exactly? How do we determine low or high fidelity occurrences? As will become apparent, defining low or high fidelity transmission and locating "occurrences of fidelity" is a tricky business, and one that has received remarkably little conceptual attention. Essentially, incorporating attribution claims about imitation and cumulative culture under a necessity-based interpretation reduces the RA to L3 and allows "skipping" the complications of fidelity entirely. Much of the empirical work derived from the RA orthodoxy, and many of its reported versions, shows exactly this take on the matter: the direct connection of imitation and cumulative culture is privileged (L3), and fidelity is conveniently backgrounded. A recent example can be observed in Caldwell and Millen's series of experiments investigating cumulative culture in the lab: despite explicit reference to the RA, this works makes no mention of fidelity,

attempting instead to test the direct connection between imitation and cumulative culture (Caldwell & Millen, 2009).

The sensation aroused by nonhuman imitation or cumulative culture (alleged) findings also strongly suggests the combined assumption of necessary conditions and human uniqueness claims. Think instead of what the situation would be under a different interpretation of the theoretical links. We could, for example, conceive of these connections instead as statistical or probabilistic ones. One likely possibility is that imitation makes it more *probable* to achieve high-fidelity transmission (and therefore cumulative culture), despite not being strictly necessary for it. I do indeed favour such probabilistic interpretation, as I think many authors ultimately do (despite their unwarranted inclination to foreground attribution claims): besides being intuitively more plausible, it has been persuasively suggested elsewhere that necessary and sufficient conditions are a rather poor instrument to think about evolutionary problems (Godfrey-Smith, 2009). Under a probabilistic/gradual/piecemeal interpretation, however, lack of imitation in nonhuman animals would make nonhuman cumulative culture only less probable, but not impossible—and consequently its potential finding not as sensational. (Notice that the starting attribution claim about imitation may or may not remain necessary, but, either way, it doesn't influence the downstream links, which are now taken to be probabilistic, in the same way.) Similarly, findings of nonhuman imitation would leave the possibility of cumulative culture open for the group, perhaps even more probable if evidence of imitation is solid. But no finding would be decisive or startling in and of itself. Lack of cumulative culture outside Homo sapiens may be then interesting and worth investigating, but no extraordinary news for the RA.

In sum, the necessity-based interpretation of the theoretical links in the RA is strongly suggested by the manner the literature has gone about investigating them, i.e. looking first and foremost for empirical evaluation of human uniqueness claims. Arguably, a probabilistic interpretation, albeit ultimately more plausible, would not have generated such heated empirical debate on the

human uniqueness of cumulative culture or imitation. Eventually, I suggest that a probabilistic interpretation of the attribution claims, *coupled* with independent investigation of the RA links, represents the best research approach. By "independent investigation" I mean that links L1 and L2 (and derived L3) should be assessed independently of the putative human uniqueness of their terms and rather on the grounds of their potential explanatory content.

Independent investigation of the theoretical links is necessary: even the probabilistic interpretation, in fact, would not be fully satisfactory per se. This is because both necessity-based and probabilistic interpretations are merely *descriptive* approaches, and as such explanatorily unsatisfying. While we may well be interested in establishing *whether* imitation is conducive or necessary to high fidelity transmission (or high fidelity to cumulative culture), providing an explanation of *why* and *how* that's the case seems ultimately more important and interesting. I now turn to illustrate more specifically what I mean by saying that descriptive relations such as necessity-based and probabilistic ones are "explanatorily unsatisfying", and to illustrate in detail how their properties and limits affect the RA.

1.5.2. Theoretical links as necessity relations: the drawback of explanatory gaps

In the previous section, we saw that there are specific reasons that the interpretation in terms of necessity has been favoured: by establishing the nesting set of necessary relations, and linking it to the attribution questions, the RA becomes more empirically compliant. However, the downside of this interpretation is that such relations are explanatorily wanting. Claims of this sort capture only a very general level of theoretical commitment; they do not offer explanatory details on *how* or *why* such states of affairs obtains (cf. for example

explicit complaints about the explanatory gap in L2 by Enquist et al., 2010; Laland, Kendal, & Kendal, 2009; Lewis & Laland, 2012).

Let me explain by contrasting explanations and necessary conditions. We can describe theoretical connections at different levels: at one level, as necessary conditions (and/or sufficient, if that's the case); at another level, they can be characterised explanatorily, with a more or less detailed theoretical apparatus that clarifies their causes, context and consequences. The two levels are not in competition, nor mutually exclusive. The explanatory level is more informationally demanding, and, unsurprisingly, more explanatorily rewarding. When we give explanatory characterisations of the links, we are calling on a completely different and much richer kind of information.

An explanatory apparatus generally specifies the types of interactions between phenomena, which may include but not be limited to necessary and sufficient conditions: a variety of evidential sources and inferential tools can be used to substantiate explanations, depending on the framework of choice (mechanistic, selectionist, teleological, historical, etc., see Friedman, 1974; Ruben, 1993). To appreciate the difference between the two levels, consider the toy case of the "triangle of fire" as an example. The fire triangle or combustion triangle is a simple model commonly presented at fire safety or firefighting courses to illustrate the necessary "ingredients" of fire, or in other words its necessary conditions:



Figure 7. Illustrating explanatorily defective necessity relations: the fire triangle

One very common application of the model is the firefighting technique known as "backfiring" (in the US), "backburning" (in Australia), or simply "controlled burn" (Burrows, 1986). When firefighters have to deal with large and rapidly spreading fires that cannot be suppressed with water or sand, they delimit a space ahead of the fire and eliminate all potential fuel within the line (usually by burning it in a controlled way). The tactic works by depriving the approaching wildfire of one of its three "ingredients": fuel (the only effectively removable one, as oxygen and heat cannot be taken under control). Thanks to its simplicity, the fire triangle is actually quite powerful: one needs to acquire only a relatively cheap piece of information (i.e. presence or absence of any of the three conditions) in order to put out a fire. Now, compare this information with one about the oxidising reaction occurring in combustion: explanations about the physics or chemistry of combustion can be spectacularly complex-at least if compared to the very elementary information provided by the fire triangle model. Nonetheless, firefighters do not need to know the details of the oxidising reaction at the molecular level, let alone the physics of combustion, to put the fire out: this information is simply redundant. The example serves to illustrate one very basic point: necessity relations can be very powerful, but informationally much cheaper than explanations.

I suspect that much of the work on RA claims has been carried out in the theoretical perspective of necessity claims, rather than in more detailed explanatory terms, exactly because the latter is much more demanding and recalcitrant to empirical treatment (Caldwell, Atkinson, & Renner, 2016). Incidentally, things do not improve under the explanatory aspect if we adopt the probabilistic interpretation of the central links. This is because probabilistic claims also, like necessary ones, *describe* rather than *explain* a specific state of affairs. In conclusion, specifying the theoretical links in terms of descriptive claims only leaves explanatory gaps in the argument; it may well turn out that the relations between terms are probabilistic or necessary ones, but an essential

part of the task of understanding and evaluating the argument calls for providing an *explanation* for such merely descriptive relations.

This feature has allowed for a worrying amount of undetected cross-talking on the subject. Explanatory gaps associated with the predominantly descriptive approach have left room for the proliferation of heterogeneous interpretations of the RA: I will examine some of them in the course of the following chapters (particularly in chapters 4 and 5) and point out their discrepancies and potential equivocations. In fact, the main objective of this work will be that of extracting and individuating the most common of the contrasting and often implicitly held interpretations of the RA. I shall argue that "the RA", beyond the minimal and consistent connection of imitation, high fidelity and cumulative culture illustrated by the diagram in Figure 1, is not a single argument at all. We would do better to think of "the RAs", for more than one explanatory path can be drawn among the essential elements of the backbone diagram. Specifically, this multiplicity of RAs derives from two features: one is the explanatory gap left by mostly descriptive approaches to theoretical links, which I have just illustrated; the other is the definitional instability of its terms "imitation", "high fidelity" and "cumulative culture". Each of these notions affords multiple definitions; depending on which one is chosen, the RA can assume rather different meanings. The rest of this work is structured according to these assumptions.

Chapters 2, 3, and 4 are devoted to examining, respectively, *imitation*, *cumulative culture* and *high fidelity*. Each chapter will provide a brief historical overview of the term (to the extent it is relevant to the RA) and an outline of the crucial conceptual difficulties associated with defining the notion and detecting the empirical phenomenon. These chapters will also strive to pinpoint and label the principal interpretations of "imitation", "cumulative culture" and "high fidelity" that have been considered relevant to the RA in the two decades following its appearance. The last chapter of the dissertation will be concerned with putting together the multifaceted picture of the RA emerging from our understanding of the indeterminacy of its central terms.

Before I move on to the task of examining in detail imitation, cumulative culture and high fidelity, however, I will briefly revise the accessory claims of the RA in the following section, and conclude the chapter with a look forward to the overall structure of the thesis.

1.6 Accessory claims

I now briefly revise what I've called the "accessory claims" of the RA. As I said earlier, I have chosen to organise these claims as a set of adjunct statements because they have less general validity than the three central claims—either fewer authors would subscribe to them being part of the RA, or they illustrate secondary, implicit, or derived aspects of it. However, these accessory statements have been at times defended very seriously, and I will occasionally return to them in various parts of the thesis. Finally, a version of claim VIII will underwrite what I think is the most plausible and informative interpretation of the RA. So it is appropriate to spend a few words on the accessory claims and clarify the related issues. Here is the list of statements as they were introduced in section 1.2:

- IV. Imitation is a uniquely human trait.
- V. Fidelity is unique to imitation.
- VI. Imitation is an adaptation for cumulative culture.
- VII. Inherent features of the social learning mechanism (fidelity in particular) are the primary factors responsible for the nature of the cumulative cultural phenomenon.

VIII. The overall argument is often restricted or thought to properly apply to a particular subset of cultural phenomena, i.e. complex material culture, technology, and skill.

I have already amply dealt with claim IV and its dwindling empirical support, so I move on to the next.

Claim V can be thought of as a "fortified" version of claim III in the minimal outline of the RA. If claim III stated that "imitation is the most or only high fidelity type of social learning", leaving some space for deviations, claim V reformulates the link in the strongest, necessity-based form. I think that stating claims in terms of uniqueness suggests the relation either as necessary, or even as jointly necessary and sufficient. But the interpretation in terms of a jointly necessary and sufficient relation should be discarded, as it would not leave room for cases of imitation *without* high fidelity—i.e. cases of low-fidelity or failed imitation—which would be clearly incorrect. A merely necessary relation instead allows for low fidelity imitation-which nonetheless, under RA assumptions, would not have the potential of generating cumulative culture. Rather, I think that a more plausible (and perhaps common) interpretation of claim V is to exclude social learning mechanisms other than imitation from being high fidelity. This is consistent also with the weaker statement in claim III: i.e. that imitation is the most high fidelity social learning mechanism. In fact, the view that emulation and other social learning mechanisms such as local or stimulus enhancement can transmit with high fidelity, perhaps under some definite conditions, is typically held as a counterargument or detracting claim for the RA (Avital & Jablonka, 2000; Caldwell & Millen, 2009; Horner & Whiten, 2005; Jablonka, 2002a; Sterelny, 2012).

The principal issue with claims III and V, however, has little to do with confounding necessary vs. necessary and sufficient relations, and much more to

do with what explanatory gaps and definitional instability. The connection between imitation and high fidelity represents a challenging and largely underrated theoretical node in the argument: how exactly would fidelity be unique to, or mostly promoted by, imitation? How does imitation implement fidelity? Again, the answers will depend largely on what exactly we mean by "high-fidelity transmission" (and by "imitation"), and on how we think fidelity gets implemented; but these questions have been largely glossed over or simply taken for granted. Chapters 4 and 5 will look in more depth into this important and unappreciated theoretical issue, which definitely deserves more than the hasty treatment it has been accorded so far.

Claim VI ("Imitation is an adaptation for cumulative culture") represents a highly controversial thesis, but it is nonetheless held in the literature, although often as a target of criticisms. (See Heyes, 2013b for critical arguments; Shea, 2009 and Tomasello, 1999 for positive ones.) The final part of this dissertation will suggest a further argument against this claim: I follow Heyes (2013b) in basing my argument on the assumption that "x can be an adaptation to y if and only if x has been selected because of y", but I hypothesise different support for my story. In brief, I will argue that there are problems in maintaining that imitation has been selected *because* it contributes to cumulative culture—the connection is, rather, indirect, and perhaps can be understood as an *exaptation* rather than an adaptation.

Claim VII ("Inherent features of the social learning mechanism are the primary factors responsible for the nature of the cumulative cultural phenomenon") captures an overall way of thinking or theoretical trend associated with, and largely promoted by, the RA. Because the discussion about the RA has focused largely on imitation, the battlefield for explaining cumulative culture has been mainly occupied with efforts to distinguish social learning mechanisms, their features and properties, obscuring alternative explanations for its appearance. The literature has been moving fast in the direction of filling this gap, and warning about the excessive and almost exclusive attention

accorded to social learning, as well as proposals of alternative explanatory factors, are nowadays endemic and even fashionable (Henrich, 2004; Lycett & Norton, 2010; Querbes, Vaesen, & Houkes, 2014; Vaesen, 2012 see chapter 4 for a brief discussion of extra factors stabilising cultural transmission). My take is that, if the goal is primarily to explain the emergence of cumulative culture, then we should definitely look beyond the boundaries of social learning mechanisms. On the other hand, social learning mechanisms are bound to be of interest to cultural evolution questions, for social learning remains undoubtedly the primary engine of culture and, importantly, the only universally and unquestionably recognised component of its definition. Understanding the contribution of social learning to cultural transmission in its different modalities remains therefore an interesting, if partial, project.

Finally, I will treat claim VIII ("The overall argument is often restricted or thought to properly apply to a particular subset of cultural phenomena, i.e. complex material culture, technology, and skill") as integral to the most interesting interpretations of the RA. The claim was placed among the accessory ones because it often doesn't figure explicitly in standard RA formulations (but cf. how the claim is considered and reviewed by Boyd, Richerson, & Henrich, 2013; Charbonneau, 2015; Hunt, 2014; Morin, 2015; Pradhan, Tennie, & van Schaik, 2012). The standard RA, having drifted into a form of conventional orthodoxy, is often presented as concerning "cumulative culture" in general, without specifying exactly what counts as "culture" (Morin, 2015). But this breadth of scope is mostly problematic, as I will argue in Chapter 3. A folk notion of "culture" would reasonably include things such as myths, religious beliefs, even norms perhaps. However, I think that, even at a first glance, such an inclusive notion of culture is not a reasonable explanandum for the RA, or at least it makes it an extremely weak and loose line of reasoning. Clearly cultural items such as myths, beliefs, and complex institutions largely depend on something other than the capacity to imitate-first of all, language. If we interpret the RA as being about imitation, as I clearly do in this work, we are

looking especially at extra-verbal means of transmission; but imitation and social learning specifically exclude linguistic transmission (Heyes, 2012a). It may be true that social learning and imitation contributed to the evolution of language (and through this to the evolution of complex cultural items), but testing this claim would give rise to an alternative explanatory project on its own. Another possibility is that social learning contributes to the transmission of complex cultural variants such as myths and norms *in parallel* to language: but it would be rather difficult to isolate its effects from the linguistic ones, and again figuring out the details of such a combined dynamic would require extensive extra work beyond the possibilities of this dissertation.

Perhaps the best strategy is therefore to highlight *delimitations* to the scope of the argument. Delimitations differ from limitations (Simon & Goes, 2013). The latter are imposed and inescapable, including all results that the study is unable to deliver due to intrinsic limitations of theoretical, methodological or empirical tools. Delimitations, by contrast, correspond to intentionally excluding some topics from the study. Intentionally restricting the scope of enquiry (in time, space, or according to other criteria) works towards a sharper understanding of particular themes, or towards preventing and neutralising questions and objections. In the specific case, for example, the RA could supply explanatory insights towards cumulative culture as a whole. However, I think it is likely to be a much more powerful explanatory tool for cumulative technology, and particularly early technology. In fact, the aforementioned delimitation conveniently configures a corresponding chronological delimitation. Extra-verbal means of transmission can be thought of as pre-verbal means of transmission, reducing the historical explanatory target of the RA to the pre-linguistic stage of hominid evolution.

Several proposed interpretations of the RA that I consider in this work do not operate with such delimitations explicitly, and will be criticised on this ground. In contrast, in this work I will intentionally focus on visuo-motor modalities of copying (i.e. on imitation and other social learning mechanisms),

explicitly excluding verbal transmission (and even auditory imitation, in fact), and also focus on the explanandum of cumulative *technology* rather than culture in general (and in particularly on early, preverbal technology). I will ultimately suggest that such intentionally narrowed version of the RA, incorporating claim VIII, is in fact our best bet to salvage its explanatory potential.

1.7. Conclusion: not one, but many ratchet arguments

This chapter has served to define the contours of my target argument, lay out its basic structure, and highlight what I think are the most pressing questions related to it and what are instead stray issues. In so doing, it has necessarily imposed a simplification. This is to some extent inevitable, because my discussion relies largely on packages of ideas that often come in an abridged form, are frequently taken for granted or indirectly reported, and because my goal has been to structure the topic for effective upcoming discussion; but no serious distortion of the central issues has occurred. I will now wrap up the discussion about the RA orthodoxy and illustrate how I intend to carry out the analysis of the argument in the rest of this work.

As I have anticipated, the RA orthodoxy implies that there is a single and univocal argument. In reality, due to the ambiguity introduced by the explanatory gaps in L1 and L2 and to the definitional instability of its basic terms "imitation", "high fidelity" and "cumulative culture", multiple competing lines of reasoning can be drawn off its basic structure (fig. 1). The main purpose of this dissertation will be to isolate and define these competing interpretations of the RA. Ultimately, this will expose some of the controversies surrounding the RA claims as apparent rather than genuine, and will allow us to assess more clearly the respective contribution of each RA's interpretation to the cumulative culture debate. One central task necessary for the identification of different RAs is the disambiguation of its basic terms. Much of the interpretative leeway in the RA depends on the existence of different definitions and accounts of, respectively, "imitation", "cumulative culture" and "fidelity". Depending on the meaning of the term one decides to plug into the RA basic scheme, different inferences and conclusions can be drawn from the theoretical links; hence getting a solid grip on what these notions mean to different authors is essential to any further theoretical work on the argument. The central chapters of this thesis are devoted to examining each notion in turn (i.e. Chapter 2 examines imitation; Chapter 3, cumulative culture; Chapter 4, high fidelity).

While the three notions face a similar problem of definitional instability, this can be ascribed to different factors in each case. For imitation and cumulative culture, phylogenetic attributional and definitional questions often appear deeply interlaced, and the historical development of these notions has a considerable conceptual bearing; so chapters 2 and 3 take something of a more pronounced historical approach to the matter. Much of the ambiguity in the notion of imitation can further be ascribed to its multidisciplinary profile; while in the case of cumulative culture the breadth of the underlying notion of culture represents a crucial factor of instability. It should also be noted that in the case of the first two notions, the difficulty posed by multiple definitions is often acknowledged, if not always subsequently met in application.

Fidelity, on the other hand, is virtually unexplored territory, and no one seems to have seriously appreciated that the notion may be ambiguous. Plenty of work relies on this notion, but very little, by contrast, tackles the question of its definition, and, importantly, its implementation in the relevant substrates at the relevant levels of analysis. So chapter 4 is similar to the previous two in addressing the definitional question for high fidelity; however, its structure diverges in other respects. As I said in section 1.4, fidelity differs from imitation and cumulative culture in its not being a biological trait of any sort, but rather an abstract property that many biological systems (inter alia) are believed to possess. This more intangible nature appears to have made the conceptualisation of fidelity more "sensitive" to the theoretical role assigned to the notion in different claims. As is evident from the diagram, fidelity is the central term of the RA and participates in L1 as well as in L2, where it connects imitation and cumulative culture. Because the definition of fidelity has tight connections with its hypothesised explanatory role on both sides, chapter 4 must deal with the explanatory gaps in both links. As a result, chapter 4 is significantly longer than its predecessors.

Laying down the different definitions of "imitation", "cumulative culture" and "high fidelity" and examining the related conceptual issues will provide most of the material needed to delineate the promised output of this work: namely, the comparison of competing interpretations of the RA. At this point, it is useful to introduce a final diagram illustrating how different meanings of the terms, when "plugged into" the RA and connected though different readings of L1 and L2, can generate substantially different logical paths and conclusions (in brief, different RAs).



Figure 8. The Abacus Diagram. Prefiguring competing interpretations of the RA.

Imagine that each of the possible meanings extracted over chapters 2, 3, and 4 represents a "bead" on the appropriate line of the following "abacus" chart. (I call it an "abacus" only because of its visual similarity.) The abacus chart is once again a redrawing of the same basic RA structure presented in Figure 1; this time, however, each term is not represented as a solid, one-piece unit, but rather as "unfolding" downward along a line that collects its different possible meanings.

Different combinations of "high fidelity" and "imitation" meanings, represented by the dotted lines connecting the "beads", will then generate different

interpretations of L1; likewise, different combinations of "high fidelity" and "cumulative culture" meanings will generate different formulations of L2. In turn, different interpretations of L1 and L2, combined into a full "RA" logical path, will generate alternative version of the RA. At this stage, the diagram necessarily presents both the "beads" and the dotted links as nondescript, empty placeholders: they will be appropriately filled in Chapter 5, drawing on the analysis of the terms and of their theoretical connections undertaken in the preceding chapters.

A final word on the information visualised in the Abacus diagram. To start with, not all the accounts of the RA terms will be entirely mutually exclusive, as is suggested by their being represented as distinct "beads"; but each of them will differ from its competitors in at least some important respect. In addition, the relative position of the beads on the downward-unfolding lines is mostly arbitrary rather than theory-laden: that is, the descending lines do *not* represent gradients or dimensions of any particular sort. It is possible, as we will see, to isolate directly "contrastive" accounts with respect to one uniform criterion; but this concerns mostly pairs rather than the full sequence of notions (this is especially the case for imitation).

Alternative interpretations of the theoretical links and of the associated explanatory gaps (when available) will be discussed in chapters 4 and 5, while the relevant RA accounts will be "recomposed" and the Abacus diagram ultimately filled out in Chapter 5. Notice that filling the explanatory gaps is not in competition with providing necessity-based or probabilistic interpretations of the argument; rather, it is part of the evidence required to decide between the two descriptive alternatives, as well as between positive and negative readings of the claims (for example, deciding between "high fidelity is necessary for cumulative culture" or "high fidelity is irrelevant to cumulative culture"). Very often, however, the explanatory content of the links will have to be extracted from indirect clues in the work of the major players in the debate.

As far as my own views are concerned, I tend to favour a soft, probabilistic interpretation of link 1 (imitation-fidelity); that is to say, imitation may be more

conducive than other social learning mechanisms, or more conducive in general, to high fidelity transmission. The stronger version of the link, which would characterise imitation as the only high-fidelity social learning mechanism, appears implausible. In link 2 (fidelity-cumulative culture), by contrast, the necessity-based interpretation stands a better chance. Strictly speaking, the RA generally presents this link as necessary. The claim may in fact be read as a special case of a broader evolutionary pattern (or even "law", some would dare say) holding that high fidelity transmission is necessary for the emergence of complex evolutionary products in all evolutionary processes, biological as well as cultural. Thus, higher things are at stake with understanding this link: should filling the explanatory gap provide reason to think that fidelity is only conducive to cumulative culture, but not necessary for it, a rather unexpected and uneasy result would have to be dealt with. Such a result would in fact contradict the general consensus about fidelity of transmission being indispensable to cumulative processes in evolution at a substrate neutral level. But of course, much will depend on the cultural interpretation of fidelity.

Finally, I tend to favour the probabilistic interpretation for link 3, which postulates a direct connection between imitation and cumulative culture, on the grounds of the probabilistic interpretation of the related link 1. Nonetheless, a further comment is in order for link 3. Normally, this is derived from the combination of link 1 and 2, incorporating the reference to fidelity. But an interesting possibility is that link 3 may be thought of as a self-standing and autonomous claim: in other words, imitation may be conducive to cumulative culture for reasons that have nothing to do with fidelity of transmission. I will briefly explore this intriguing possibility in the final chapter.

Having thus concluded the discussion of the basic outline of the RA, and of how this will structure the upcoming work, I now turn to presenting the terms of the argument in turn in the next three chapters: namely, *imitation, cumulative culture*, and *high fidelity*.

CHAPTER 2. What is Imitation?

2.1. Introduction

The first section of this chapter (2.1) explains why defining imitation is important for the RA, the challenges of the task, and how I intend to carry it out.

Section 2.2 outlines a short history of the notion and highlights two possible meanings of "imitation" associated with its historical interpretation (imitation as "all copying" vs. imitation as a specific kind of copying). It also introduces modern social learning studies and imitation's place in this context.

Section 2.3 discusses imitation in the framework of social learning in greater detail and presents two examples of contrastive definitions of "imitation", pointing out the most important definitional criteria adopted for imitation and their shortcomings.

Section 2.4 looks at another source of ambiguity, isolating two contrasting views on social learning, and correspondingly on imitation (imitation-mechanism vs. imitation-channel). The section suggests that the two should be kept carefully distinct on the ground that channels and mechanisms are two fundamentally incommensurable notions (albeit with important connections).

Section 2.5 looks in particular at the relationship between mechanisms and channels, at the problems deriving from failing to distinguish clearly between these two notions, and at solutions to such problems.

Section 2.6 concludes the chapter by formulating accounts of imitation relevant to the RA from the initial set of ambiguous ones. I will argue that imitation should be considered first and foremost a social learning mechanism sourcing information from behavioural morphology to potentially produce matching behaviour. A few extra features of imitation that fit this definition are also discussed and selected in this final section.

2.2. Defining imitation and its place in social learning

"There is no **absolute** meaning of "imitation" that we might waste time debating. Each investigator will be well advised to define such terms at the time of each new application of them." (Whiten & Ham, 1992:249)

The success and strength of an argument depends not only on its logical articulation, but also on how conveniently it defines its basic terms and notions. A primary reason to take the definition of imitation seriously is therefore that what we *mean* by "imitation" can substantially change the truth-value of the claims made in the RA. A secondary reason is that the current state of the art of definitional endeavours for "imitation" is a fractious matter. There is more than one possible way of defining "imitation", and, even more, definitional strategies appear to follow a disparate, incongruous array of criteria (Hurley & Chater, 2005; Zentall & Galef, 1988). This conceptual fragmentation, while causing palpable concern among imitation and social learning specialists, tends to go ignored in outer research fields, which nonetheless place high theoretical stakes on imitation (cf. Acerbi, Tennie, & Nunn, 2011; Caldwell & Millen, 2009; Shipton, 2010; Sterelny, 2003). The widespread assumption appears to be that we all

know, after all, what imitation is about: it is *copying*, isn't it? Unfortunately, things are not so simple. The available accounts of imitation suffer from subtle but significant variation, and the taxonomic literature on social learning (which aims primarily at defining, detecting and classifying different mechanisms), due to its stratification and convolution, struggles to reach a meaningful integration with the wider scope theoretical literature, especially that which incorporates imitation in evolutionary arguments (like the RA does).

To some extent, the under-determination afforded by extremely broad notions might have contributed to the fortunes of the RA. Statements and definitions of general scope are often in the position to generate wider agreement. Many aspects of the RA tap into this "everyone can have their say" feature: the argument appears prima facie intuitively plausible, involves everyday notions and somehow eludes their overly strict definition. This has afforded it broad consensus and diffusion. On the downside, it has allowed the proliferation of a baffling variety of interpretations, with the argument appearing alternately obvious, almost trivially true, or grossly implausible.³ With the appropriate delimitations and sharper definition of its terms, however, the RA may still have genuine potential for shedding light on cultural evolution dynamics: but it is necessary to determine whether apparently opposed views are genuinely inconsistent, and whether apparently complementary views are actually complementary, and this task rest primarily in the disambiguation of its terms.

Ambiguity in the definition of "imitation" has various sources, which I will group into three broad categories. First, it can be imputed to *historical* factors: research on imitation has spanned over a century (and general interest in the

³ Among the most dumbfounding examples of inconsistent interpretations of RA's claims: "imitation is necessary for cumulative culture" (Richerson & Boyd, 2005; Tomasello et al., 1993) vs. "cumulative culture doesn't need imitation"(Avital & Jablonka, 2000); "high fidelity is the key to cumulative culture" (Lewis & Laland, 2012) vs. "cumulative culture happens without high fidelity social learning" (Claidière & Sperber, 2010; Morin, 2015; Sperber, 1996).

subject can be found much earlier, Galef, 1988; 1992). Over time the focus and motivation for interest has shifted, new theoretical and empirical considerations have been added to the picture, and definitions of "imitation" have fluctuated accordingly. The historical source of ambiguity will be examined particularly in sections 2.2 and 2.3, and will deliver a very broad, crosscut distinction in the available definitions of imitation. On one hand, "imitation" will appear as a very general term to indicate all copying behaviours; this out-dated but stubbornly die-hard account will stand in contrast to imitation as *specific copying*. It is fair to say that in contemporary literature that is reasonably specialised with the RA, imitation is universally considered a very specific type of copying mechanism (just how specific, however, is a matter of considerable debate, which will constitute the main matter of the third, intra-disciplinary source of ambiguity listed below). The distinction between *imitation as all copying* and *imitation as* specific copying is nonetheless worth rehearsing, for even in contemporary studies, "imitation" is at times used in its old, unsophisticated sense of all copying, whether ingenuously or not. After I have clarified the historical ambiguity and sketched the basic distinction between imitation as all copying and as a specific kind of copying, the second meaning will effectively drop out of the picture. Only imitation as all copying will make its way into the Abacus scheme, while imitation as specific copying will serve as a working umbrellaconcept, from which further RA-relevant distinctions will be drawn.

Secondly, *interdisciplinary* factors are to be considered: imitation is a "hot topic" in at least two clearly distinct disciplines—namely, psychology and cultural evolution theory—in addition to having a place in a number of other connected scholarly sectors possessing in turn their own partial paradigms and agendas (e.g. comparative, evolutionary, developmental and social psychology; cognitive science and neuroscience; behavioural sciences and artificial intelligence; archaeology, anthropology, sociology, human evolution, behavioural ecology, and arguably many more, cf. Heyes, 2017; Hoppitt & Laland, 2013; Mesoudi, 2009). The different goals, interests and historical

developments of such disciplines have emphasised incongruous aspects of imitation, producing slight (and sometimes not so slight) variations in its definition. Crucially, I will argue that the inherent interdisciplinarity of imitation causes an important ambiguity between accounts of *imitation as mechanism* and *imitation as channel*. This is in effect a broader problem that extends to the global topic of social learning, with much literature on the RA (and more generally on cultural evolution) often being equivocal about whether it is about social learning *mechanisms* or social learning *channels* (StereIny, 2009). While this source of ambiguity is theoretically crucial, again it will not give rise to separate, competing account of imitation to be plugged back on in the Abacus scheme. Rather, it will serve illustrate a very widespread assumption about how different social learning mechanisms are supposed to contribute to different social learning channels. The interdisciplinary ambiguity will be discussed in sections 2.4 and 2.5 of the present chapter.

Finally, the conceptual landscape for imitation appears fragmented and polysemous even within the same field of studies and time span, due to explicitly divergent proposals for its definition within social learning studies. This last source of ambiguity may thus be named intra-disciplinary. As mentioned above, imitation started out historically having a very general, all-encompassing meaning of "all copying"; this meaning got progressively narrowed down in some form of "specific copying", but not always in the same direction or according to the same rationales. While the general trend is to categorise mechanisms according to their input condition (i.e. what kind of information about a social scene they allow the subject to intake), sometimes other considerations intervene in the classification (for example, level of intentional control, the cognitive machinery supposedly deployed, or motivational factors). As a result, the current social learning literature understands significantly noncoextensive referents as "authentic imitation" (cf. Gerrans, 2013). The intradisciplinary definitions will be discussed in section 2.7, and will generally constitute authentically competing account: "beads" in the Abacus scheme.

While the fundamental aim of the chapter remains that of delineating as sharply as possible the several competing accounts of imitation that can be (and have been) plugged into the RA, I should point out that my approach to the problem is somewhat unlike the one regularly adopted in the literature—and may appear circuitous at first sight. As a rule, authors who show an appreciation of the ambiguity of "imitation" undergo complex apologia in their opening pages (e.g. Bates & Byrne, 2010; Byrne, 2003; Call & Carpenter, 2002; Zentall, 2006). They note, as I have done, that imitation is a vast subject, one that has sparked interest in such disparate timeframes and fields that it is hard to talk about it in a unified and coherent manner. In virtually all cases, however, what follows these skirmishes is a *stipulative* or working definition of imitation—that is (as suggested by Whiten and Ham's opening quote to this chapter), an individually rephrased characterisation of imitation that fits the scope and purposes of the job at hand. Stipulative definitions are certainly a convenient and intellectually transparent shortcut; they can sooth vagueness in the immediacy; but alas they do not add to the clarity of the field as a whole.

My strategy will be different—for my goal is not to isolate a working definition of "true" imitation, but rather to show that extant working definitions are indeed seriously disparate. Essentially, I will aim to provide the reader with coordinates to navigate categorisations of social learning first, and definitions of imitation as a result: the incongruities and conceptual hurdles that characterise the topic of imitation in fact cannot be understood unless in the broader context of the social learning debate.

The problem of categorising social learning affects definitions of "imitation" in several ways. A primary connection rests on historical grounds. "Social learning" is a reasonably new term. A "social learning field of studies," as we intend it today, was non-existent only a few decades ago; but several aspects of the same phenomenon had attracted scientific interest for much longer under the general label of "imitation" (Galef, 1988; Heyes, 2017). Secondarily, "imitation" has often been defined *via contrast* to other social learning

mechanisms, initially appearing as "spurious" cases of imitation. Two examples of this contrastive categorisation strategy, involving the two well-recognised social learning mechanisms of emulation and local enhancement, are discussed in section 2.4. Contrastive definitions, and in particularly the one that sees imitation rivalling emulation, have had a significant impact on the understanding of imitation as a special social learning mechanisms, and represent perhaps the most common ways of thinking about imitation, especially in the RA context (see Acerbi, Tennie, & Nunn, 2011; Caldwell & Millen, 2009; Tennie, Call, & Tomasello, 2009). Finally, because in several versions of the RA imitation is considered fitter for yielding high fidelity transmission than other social learning mechanisms, a comprehensive overview of the field is important to gain a better understanding of the reasons of this contrast. With these provisos in place, I now turn to discuss the various sources of ambiguity in imitation.

2.3. Imitation: brief history of an idea

Interest in imitation is not new to the RA. Studies of imitation date back at least to the 1800s in modern scientific terms, but the fascination exerted by *mimesis* on intellectual endeavour can be traced back much further, right to the origins of philosophy. Plato and Aristotle debated *mimesis*, the Greek word for imitation, with Plato seeing mimesis as a decayed, "twice removed" copy from the world of ideas, and Aristotle seeing it as a fundamental disposition of human beings to represent reality (Baktir, 2003; Belfiore, 1984; Else, 1958).

Luckily, we need not concern ourselves with the notion pondered by ancient philosophers. This is probably as broad as the everyday, folk concept of "imitation", which may legitimately describe phenomena as diverse as copying a painting, a movement or an attitude, performing the parody of a character, following a fashion trend or plagiarising a tune. Nonetheless, what unites these disparate referents does have a bearing on the more technical and modern notion of imitation, which is the subject of this chapter. This element may be identified with the "mimetic effect", broadly denoting the duplication of an original item into a similar one. The mimetic effect has always constituted the crux of the interest taken in imitation, and contemporary scholarly studies follow suit, albeit for different reasons, to different outcomes, and (sometimes) on dubious grounds.

In contrast with the folk notion, modern-day scholarly notions of imitation are generally restricted to visuo-motor imitation of a conspecific—that is, the capacity of acquiring motor behaviour from its visual observation, in an intraspecific context (rather than, say, copying music, paintings, or dress codes; and rather than in interspecific context). Visuo-motor imitation is clearly the phenomenon on the radar of modern studies, as testified by the classic and omnipresent Thorndike quote that depicts imitation as "learning an action from seeing it done" (Thorndike, 1898; 1911). The standard demonstrator/observer terminology, ubiquitous in social learning debates, suggests the same diagnosis (Byrne, 1999; Hoppitt & Laland, 2013). Even though restricted in principle to visuo-motor imitation, modern definitions display various forms of ambiguity, the first of which is examined in this section.

It is useful to draw a rough chronological partition in modern imitation studies. We can identify two main distinct stages, corresponding (again, rather crudely but informatively) to two distinct perspectives and above all to two distinct usages of the term "imitation". The first stage may be identified with the above-mentioned early investigations of nonhuman cognition, in the decades straddling the 19th and 20th centuries. The second dates back to the last decades of the 20th century, when imitation underwent a revival of interest, marking the beginning of "social learning" studies. Symptoms of this trend were emerging in the 1950s, but they definitely gained full traction through the 1980s (Galef, 1988; 1990).

Early studies focused predominantly on imitation's cognitive significance. There had been a time when imitation seemed nothing but a cheap trick, typical of "inferior" primates and thereby attesting the (alleged) superiority of human rational, mindful behaviour (as the verb "to ape" strongly suggests, see Byrne & Russon, 1998). But suddenly researchers seemed struck by the wonderful cognitive puzzle posed by the capacity for "learning to do an action, from seeing it done". The primary preoccupation of early investigators such as C.J. Romanes, C.L. Morgan or E.L. Thorndike in (Galef, 1988; 2013) was to explain how externally acquired representations of behaviour can guide production of a similar behaviour in the observer—perhaps especially puzzling when the latter cannot see herself performing the action, as in imitating facial expressions (Meltzoff, 1988; Meltzoff & Moore, 1977, 1997). The cognitive underpinnings of such a faculty appeared to be vastly sophisticated, and as a consequence, ca va sans dire, probably uniquely human. This early association of imitation with cognitive complexity and humanity left a long-lasting mark on the field. In more recent times, imitation became associated with high-level cognitive skills, such as mindreading or perspective taking, which in turn depended on a certain way of portraying the mechanisms underlying action reproduction.⁴

The point that I want to highlight here is that, in this earlier stage, "imitation" denoted in general *all* mimetic behaviour—that is, a broad capacity to reproduce behaviour through observation. The word "imitation" was an umbrella term to define any behaviour with mimetic effects. It may be said, in a

⁴ If we assume that the only way to imitate a model is to translate a representation of her point of view into a representation of the observer's point of view and vice versa, then we are forced to assume that the agent is capable of representing (and translating) representations- requiring theory of mind. Although this underlying hunch proved to be wrong (Sterelny, 2003), the exact mechanisms allowing action matching are still poorly understood.

retrospective slogan, that *all* social learning (or, better, all copying) was simply called "imitation" (Byrne, 1999; Galef, 1990).⁵

This usage contrasts sharply with the one found in the second wave of contemporary studies. In the second half of the 20th century, social learning studies started to grow as a free-standing field and to attract increasing attention (Galef, 1988). The introduction of the expression "social learning" can in fact be traced back to this later phase, with imitation gradually assuming the contours of a *specific* social learning mechanism, a subset of a wider category. Over this period, different social learning mechanisms started to be carved out from the apparently uniform phenomenon of "copying". The crucial insight that stands at the basis of this "social learning revolution" is the idea that several factors (other than chance) can generate behavioural similarity between individuals—in other words, that "copying" can be achieved in various and importantly different ways (Tomasello, 1996). Under this idea, the mimetic effect remains central, but different routes to it are identified and named as different social learning mechanisms.

The centrality accorded to the mimetic effect reflects a different interest in this second period compared to the previous phase. It has long be recognised that interest in imitation is fundamentally articulated over two foci: culture and cognition (Galef, 1988; Heyes, 1993; Zentall, 2001). If cognitive themes such as the role and use of behavioural representations characterised the early modern investigative efforts (a topic that is often described as the "correspondence problem" in modern jargon), the second wave of studies is definitely more concerned with the evolutionary potential of social learning. The nascent field of cultural evolution rapidly incorporated social learning studies, attributing to the phenomenon the capacity to function as an inheritance channel for cultural

⁵ In reality, Thorndike and other theorists mentioned above had already begun to differentiate between simpler and more complex ways of acquiring information from conspecifics—of learning socially, in effect (Galef, 1990). But the idea of "social learning" had definitely created no traction in the wider audience, and imitation was mostly intended simply as copying an observed behaviour.

phenotypes. The RA epitomises the shift from cognitive to cultural preoccupations, and has certainly contributed significantly to it.

Some other factors may have influenced this trend. One such factor can be identified with so-called "memetics". Memetics was one of the first attempts to apply evolutionary ideas to culture, albeit with an impact that may be viewed ultimately as more significant to pop culture than scientific circles (Benitez-Bribiesca, 2001; Godfrey-Smith, 2009; Mesoudi, 2009; Sterelny, 2007). Essentially, memetics sees cultural traits as cogno-viruses, selfish cultural particles that compete for fitness and spread through (sometimes at the expenses of) their carriers (Blackmore, 1999; Dawkins, 1976; Dennett, 1993). The idea is due to an extension of Dawkins' replicators into the cultural domain, but it has been criticised on various grounds and is nowadays a dwindling research program.⁶ Nonetheless, it is fair to say that memetics contributed to the dissemination of the debate on cultural evolution, and perhaps even to the fostering of ratchet-like ideas, given its emphasis on memes' fidelity of transmission and the very etymology of the term "meme".

But let us focus on the meaning of "imitation" during this historical transition. The social learning era marks an important shift. Imitation is no longer *any* reproduction or copying of behaviour; it indicates instead *a particular way* of obtaining the mimetic effect. So the meaning of "imitation" mutated historically from covering *all copying*, to covering a narrower set of cases of *copying of a certain kind*. As far as the RA is concerned, this form of ambiguity has often, but not always, limited significance. Even in contemporary studies, "imitation" is still sometimes used in its old, unsophisticated sense of *all copying* (for a critical remark, see Heyes, 2017; for possibly intentional unsophisticated

⁶ Among the most serious criticisms there is the accusation the memetics overloads the explanation of evolutionary processes with an unnecessary agential component (Godfrey-Smith, 2009) and that its explanatory benefits are redundant: local psychological dispositions can explain memes' spread and distribution, and it is not clear that appealing to memes' fitness accrues any real advantage in identifying a general underlying rule for diffusion (StereIny, 2006; Wimsatt, 1999).
usage, see Morin, 2015). But clearly "imitation" as *all copying* makes the RA either uninformative or paradoxical. It makes it paradoxical because we lack any term of comparison, in case the claims are read as singling out imitation as opposed to *other* mechanisms (what *other* mechanisms are there to be compared for their fidelity potential, if imitation describes indistinctly all mimetic behaviour?). It makes it uninformative (or tautological) because under at least certain interpretations of fidelity and culture, there is no way of telling these two notions apart from the mimetic effect, i.e. the transmission of behavioural phenotypes, which is also the fundamental criterion for "copying" (thus the argument would turn in a pretty unexciting case of the following assertion: "Transmission is necessary for transmission, which is necessary for transmission"). This last point will become clearer over the next two chapters, as it depends on the analysis of the other two terms ("fidelity" and "culture") carried out in the next chapters.

Two more remarks should be made about the new meaning of "imitation" emerging from the social learning revolution. First, imitation may now be a particular way of copying, rather than indiscriminate copying—but it is *copying* nonetheless. The mimetic effect is still central, and is, in fact, seen as the crowning definitional criterion for the whole category of social learning (for cultural inheritance depends on it—more on this in next section). This is apparent, for instance, in the classification reported below, which is one of the earliest, most influential and comprehensive, and regularly quoted as a source for the definition of imitation (Whiten & Ham, 1992).





Secondly, although the discrimination of different mechanisms is a coherent trend, it proceeds in a somewhat fragmentary manner, with perplexing results. Often the objective is to tell imitation apart from other mechanisms, with imitation being widely perceived as something of an *enfant prodige* or "Holy Grail" in the social learning family (Galef, 2014)—perhaps as a result of the

cognitive sophistication attributed by earlier studies, which was further magnified by the RA. But the criteria adopted to carry out the discriminatory task are rather mixed and often underdetermined.⁷ Inconsistency is not a prerogative of imitation, but extends also to the other mechanisms: many of the distinctions and groupings illustrated in fig. 1 have now become obsolete, having been overwritten by new and incongruous ones (cf. (Hoppitt & Laland, 2013).

Overall, I will suggest in the next section that the heterogeneous wording of criteria does not always represent a deep problem. It is still instructive, however, to examine in detail two cases of the "discovery" of new social learning mechanisms, distinguished from imitation. I do this in the next section.

2.4. Historical ambiguity: all copying vs. a specific way of copying

2.4.1. Imitation vs. local enhancement

A paradigmatic case of a contrastive definition of "imitation" is the distinction drawn between the imitation and local enhancement (and stimulus enhancement, a connected mechanism, see Galef, 1988b; Heyes, 1994a), which is associated with the classic example of blue tits. In the early 1920s, British

⁷ For example, Heyes famously defines imitation as "learning *through* behaviour, *about* behaviour", as opposed to social learning which is characterised as "learning *through* behaviour, *about* the environment" (Heyes, 1993). Another influential definition by Heyes is "copying the *topography* of observed body movement" (Ray & Heyes, 2011), which has often been rephrased as "copying the *morphology* of behaviour" (Nehaniv & Dautenhahn, 2001; Voelkl & Huber, 2000; 2007). Elsewhere, imitation is defined simply as "copying actions" or strategies or techniques (in general, as opposed to emulation, cf. Acerbi et al., 2011; Caldwell & Millen, 2008a, 2008b, 2009). Tomasello, along with Byrne and others, requires that for *true* imitation to occur and rule out other social learning mechanisms, the action copied should be *novel* to the observer (Byrne & Tomasello, 1995; Tomasello, Kruger, & Ratner, 1993).

birds of this species started to display a curiously mischievous behaviour. The birds were observed looting milk from glass bottles left on people's doorsteps, by prising open or pecking through the aluminium foil cap. The first-known record of this well-known adaptation was taken in Swaythling, Southampton (Lefebvre, 2013; Lefebvre & Bouchard, 2003), but the behaviour spread quickly across the whole country (and was picked up by other bird species as well, cf. Hinde & Fisher, 1951). The speed and pattern of diffusion did not appear consistent with independent innovation; the behaviour must have been learnt by observing other birds opening the bottles. The obvious interpretation, according to the folk and historical, pre-social learning notion of imitation, is that birds had been imitating each other; however, the phenomenon was described by some commentators as a case of local enhancement, in explicit contrast with imitation (Galef, 1988; Sherry & Galef, 1984, 1990; Spence, 1937).⁸

The difference between local enhancement and imitation has been characterised in various ways, but one of the clearest appeals is to the *source of information* attended by the observer. In local enhancement, the naïve individual is simply drawn to a location (milk bottles) by the presence of a conspecific, and subsequently works out the target behaviour (bottle opening) through individual trial-and-error attempts. The source of information used by the birds is simply a new *location* (i.e. milk bottles), highlighted by the presence of conspecifics (which represent the social stimulus). Because there are only a few successful bottle-opening strategies for blue tits (either pecking or prising open the tin foil), the resulting behaviour appears similar to that of their predecessors (Brown, 2017). Thus the mimetic effect is achieved but, according to the new interpretation, this is *not* imitation. To count as imitation, instead,

⁸ Stimulus enhancement is a related category of social learning, consisting in the observer learning to direct action to an object (stimulus) rather than a location highlighted by demonstrator's behaviour. It is in fact unclear whether the distinction between the two mechanisms is significant, and often "stimulus enhancement" is just a more general term to indicate enhancement, whether of location or object (Bennet G Galef, 1988; Zentall, 2006).

the blue tits should have directly used the behaviour of their conspecifics as a source of information to acquire bottle-opening.

In other cases, the distinction between imitation and local enhancement has been cast in terms of the role played by the demonstrator in learning. In imitation the role of demonstrator is direct; in local enhancement it is indirect, as it merely acts as a salient stimulus to guide behaviour to a location, where individual learning and innovation are performed (cf. for example Whiten & Ham, 1992; Zentall, 2001). Another way of discriminating imitation and local (or stimulus) enhancement appears to be the *reliability* of the mimetic effect. In local enhancement, it appears to "happen" (as a mere contingency), because behavioural options are constrained and channelled by the environment. If more options had been available, either for the nature of the task (imagine bottles with a more complex opening procedure, impervious to individual innovation) or for the equipment of the observer (imagine a larger bird with stronger talons that could rip open the foil or even raise, drop and smash the bottle to gain access to milk), we may not have observed reproduction of behaviour. If learning were authentically imitative, instead, the mimetic effect would be obtained more reliably, because the information used to guide novel behaviour is the original behaviour itself (cf. Brown, 2017; Hoppitt & Laland, 2013).

Ultimately, at least in the particular case of local enhancement, I do not think that the heterogeneity of criteria is a major problem. The descriptions appear to be largely coextensive, and pick out the same relevant phenomenon in the world. The heterogeneity is in fact only superficial: the second and third criteria (direct/indirect role of demonstrator; reliable vs. unreliable transmission) may be easily rephrased as logical consequences of the first (local vs. behavioural information source). If birds direct attention primarily to location, then behaviour *must* be a secondary attention focus (as the nature of attention is selective, Johnson & Proctor, 2004) and therefore an indirect factor in learning. If information comes primarily from location and not from behaviour, birds have to innovate individually, which introduces variability in the

62

behavioural response. Transmission of the same behaviour thus appears less likely, less reliable, owing similarity mostly to environmental canalisation. In sum, the source of information attended by the observer appears to be the best criterion to discriminate between imitation and enhancement.

In general, I think that source of information represents a superior and perhaps truly unifying criterion for the classification of social learning and imitation, although it needs to be conceptualised more precisely than what has been undertaken so far. In effect, the criterion recurs consistently in the literature under different names, for example as "content" or "what learning is *about*" (Heyes, 1993); or as "what is learned" (Hoppitt & Laland, 2013).

I now discuss another example of social learning categorisation, the emulation vs. imitation distinction. This example shows exactly how even the "source of information" criterion has not been applied consistently and how, perhaps as a result of insufficient conceptualisation, it has produced conflicting classifications of the phenomenon of emulation.

2.4.2. Imitation vs. emulation

Tomasello (1990) is credited with having introduced the term "emulation" into the contemporary social learning debate (borrowing it from Wood, 1988). Tomasello's observations stemmed from the description of tool usage learning in chimpanzees. In his experiment, young chimpanzees had the opportunity to observe conspecifics cracking nuts with hammer stones, deploying various techniques. The subjects did show a tendency to pick up the behaviour, but the way they employed the stone utensils appeared somewhat unspecific. Youngsters seemed rather to learn the association of tool, open nut and available food, with their attention being drawn to manipulating the stone and then to experiment with it, but without really reproducing the most complex techniques performed by the trained demonstrators. According to Tomasello, it seemed inappropriate to describe the behaviour as "imitation", because the chimpanzee had typically learned to "reproduce the completed goal (e.g. open a nut) by whatever means it may devise, including using the tool in an unspecified manner" (Tomasello 1990: 284).

The idea behind this distinction has since undergone a curious and rather perplexing fate, presumably due to an unfortunate initial formulation (as Tomasello himself recognises, see below). To date, I think it is fair to say that the by far most common definition of *emulation* is "reproducing outcome, goal, or end results of a behaviour", in contrast to *imitation*, which is "reproducing outcome *and* actions, strategy or technique" (cf. Caldwell & Millen, 2009). However, Tomasello claims that this is not what he meant; he intended rather to capture cases when "the learner observes and understands a *change of state* in the world produced by the manipulations of another" and then reproduces it (Tomasello 1996: 321, my italics). Tomasello's idea was that what chimps were learning was not the action of cracking, which would have justified speaking of imitation, but rather the fact that stones can crack nuts. I may observe you displacing a log to extract succulent larvae; but all I may learn is that the log can be rolled away, revealing a food cache.

The difference in interpretations is substantial. Under Tomasello's intended formulation, emulation is a case of learning about the *environment* and potential changes occurring therein, distinguished from imitation because imitation entails learning about behaviour. But under the interpretation that has later been ascribed to Whiten (Heyes, 2013b; Tomasello, 1996), and that currently appears most widespread, emulation is a case of learning about *behaviour*, only about a specific aspect of behaviour: its results. It is thus contrasted with imitation in that the latter is learning about the totality of behaviour, its outcome as well as the actions or strategies performed to achieve it. The clash of interpretations has been remarked upon on various occasions, with a recent intervention by Heyes clarifying (supposedly) that "at its valuable

root, the imitation-emulation distinction draws attention to *what* is copied by the observer of instrumental task performance: the *body* movements (imitation) or the *object* movements (emulation) or both.... But this *what* distinction is sometimes confounded with a *how much* distinction: between the copying the *sequence* and the copying of an *endpoint*" (Heyes 2013b: 323, some Italics are mine).⁹

Contrary to Heyes, I think that both distinctions are valuable, for they are diagnostic of two important things. One is the broad convergence towards factors relating to the *input phase* of learning (see below) and particularly the source of information, as the general criterion of classification for social learning. In particular, imitation emerges from these classifications as copying due to attending behaviour directly. Again, this way of defining "imitation" may need to be qualified in various ways, but for the moment it represents a reasonable guide.

The other is the idea that social scenes (i.e. scenes involving other individuals, mostly intended as conspecifics) are rich informational sources. There's information about the other organism, about the surrounding environment, and about their interactions. Most importantly, there is information stemming from behaviour itself, and this source also appears multifaceted and uneven. At least one popular definition of "emulation" is drawn by distinguishing different components in the source of information constituted by behaviour: emulation is interpreted as copying only the outcome of behaviour in its entirety. I will leave the theme of behaviour as a rich informational source to the last section; let me now say more on the input and output phases of learning.

⁹ Some authors have identified up to four supposedly discrete meanings of "emulation" (Huang & Charman, 2005), which illustrates the extent of the disagreement and terminological fuzziness.

2.4.3. Input or output?

This "input-output" terminology is not current in social learning studies (although it is more frequent in Al-oriented studies, cf. Morikawa, Agarwal, Elkan, & Cottrell, 2001; Nehaniv & Dautenhahn, 2007). However, I think it's convenient for capturing a broad, standard way of conceptualising learning events, without particular commitment to their actual implementing details. The scheme below thus exemplifies the "folk representation" of a prototypical learning event that is deeply ingrained in our thinking, and, I believe, tacitly underlies all the classifications examined so far.



Figure 10. Schematic conceptualization of a prototypical social learning event.

What should be noted at this stage of discussion is simply the possibility of identifying three temporal phases in a social learning event. The first consists in the exposure to a model engaging in a specific behaviour (in this case, a black rat

peeling scales off a pine cone).¹⁰ I refer to the first as *input* stage. The third consists in the production of the behavioural response affected by (the acquisition of information in) the input phase, which is assumed to be mimetic in the drawing—in accordance to the classifications revised so far. I refer to this as *output* stage. In between I add an *intermediate* stage that is intended to capture the time interval wherein a range of events (internal and external to the organism) may occur before the behavioural output.

The classifications of social learning presented so far rely on this model in taking the *output* phase as the *lumping definitional criterion* for the whole family of social learning (cf. Whiten and Ham's table above); and in taking the *input* phase (or better: elements associated with it, such as source of information) as *splitting definitional criteria* to differentiate among mechanisms.

This therefore appears to accurately describe efforts of social learning classification and identification discussed so far. Yet it invites a question. What if we knew for sure that an observer has attended and acquired information stemming from morphology of behaviour (which seems to be the definitional criterion for imitation) but no mimetic effect was observed? Would that count as imitation, or not? Imagine that the observer rat has effectively attended and acquired information from its model's actions on the pine cone (and we know for sure it has). However, it may have a previously established aversion to pine cones, and would never approach one independently; or it might never be presented with a pine cone again in the future, and as a consequence it would never get to perform the observed scale-peeling actions (more on this later). It might seem like this case *should* qualify as imitation, insofar as it ticks the boxes for the input condition; but perhaps it *should not* count as imitation, as the

¹⁰ This behaviour has effectively been observed in black rats (*Rattus rattus*) dwelling in pine forests around Jerusalem, and it is considered a classic example of social learning: naïve rats do not display the behaviour unless exposed to a skilled conspecific. In addition, it has been noted that the scales removal technique is impervious to innovation, because the peeling is effective only if initiated at a specific location: the top of the pine cone (Aisner & Terkel, 1992; Terkel, 1995; Zohar & Terkel, 1991).

output condition box remains in effect unticked (no matching behaviour observed). This conundrum depends on a further ambiguity in the definitions of "social learning" and "imitation", which I now turn to discuss.

2.5. From copying to learning: Perspectives from psychology and cultural evolution

Consider the following scenario: a young elephant seal Y observes a dominant alpha male A aggressively attacking and badly injuring a beta male, B, for coming too close. As a result of this observation, Y learns to avoid A—presumably having learnt that A is dangerous. In this example, Y has undeniably attended to a *social* stimulus (A beating up B), and has equally indisputably *learnt* from it. Shall we consider this an instance of social learning, then? Under the definition of social learning emerging from the previous section, the answer is no. The definition took mimetic effects (producing similar behaviour) as definitional—in other words, only "copying" counted as social learning. If the youngster reacts by steering clear of the alpha male, it is pretty straightforward he hasn't acquired any "similar behaviour"—he hasn't copied A—in fact, Y has learnt a completely different kind of behaviour (fleeing) from the observed one (aggressive display). No mimetic effect whatsoever is achieved; no social learning is observed.

The problem with the paradoxical situation illustrated above is due to the existence of two separate notions of social learning—and, as a consequence, of two corresponding definitions of "imitation". Under the first definition, "social learning" means to acquire the same behaviour as the one observed; under the second definition, it means acquiring information that can affect behaviour, through a social stimulus. In the second definition, whether the acquisition of social information leads to the same behaviour as the one observed or to another response (e.g. fleeing after having observed an act of aggression) is

irrelevant. In this characterisation, the mimetic effect recedes and the concept of learning comes to the fore. Social learning is first and foremost an instance of *learning*, defined broadly as a relatively long-lived change in behaviour due to experience (Seel, 2012); what makes it *social* is the nature of the stimulus observed, as transparently illustrated by Heyes in a recent paper (Heyes, 2012a), and implicitly reflected by earlier social learning categorisations (Zentall, 1996, 2001). The definition of "imitation" that emerges from this characterisation of social learning is thus best understood as *learning*, rather than *copying*.

Since I am going to maintain that the second characterisation is the correct one and ultimately superior for the purpose of discussing the RA, it is useful to note explicitly some of its important features. I shall mention three.

First, social learning is commonly contrasted with symbolic language, communication and learning-from-teaching. This is because in such cases the behaviour of the model (the source of the learned information) is adapted or intended to communicate information to a learner. In other words, information acquired through social learning is best defined in terms of *cues* rather than signals, where the latter are information channels shaped by natural selection and the former are not (Saleh, Scott, Bryning, & Chittka, 2007). Thus, social learning is often thought of as learning autonomously through observation or eavesdropping, rather than learning through the active, intentional engagement of a conspecific in passing on information (Heyes 2012a, 2013b). Notice that if we take seriously this definition of social learning, and consider imitation a specific subcategory of it (rather than in its all-purpose meaning of all-copying), our hypotheses about the transmission potential of imitation must be limited in various ways. As I have noted in chapter 1, it is in principle hard if not entirely intractable to distinguish the differential contribution to cultural transmission of social learning mechanisms and more sophisticated tools like language, in organisms that rely on both. Thus this depiction of social learning and imitation supports the idea we should constrain our hypotheses about imitation to an early timeframe.

69

Secondly, social learning is most naturally contrasted with individual learning, with the latter being defined as learning about the environment or other agents that is not mediated or influenced by social stimuli—this is in effect the main rationale of the definition. The categories of so-called individual learning most often juxtaposed with social learning are trial-and-error learning and innovation (invention) of a new behaviour. (One caveat here: as noted before, Heyes has recently pointed out that all learning ultimately takes place at the level of the individual, so social vs. *asocial* learning is probably a better terminology, see Heyes 2012a. I think that this suggestion is sensible and henceforth endorse Heyes terminology.)¹¹

Finally, note that this second and more inclusive characterisation (freed from definitional mimetic effects) may be better suited to highlight other important evolutionary functions of social learning besides that of working as a cultural inheritance channel (Jablonka & Lamb, 2005). On a very broad level, social learning represents a powerful adaptation: organisms that can predict, avoid, and manipulate events on the basis of previous experiences are immensely better placed than those who cannot, no matter what kind of experiences we're considering (social or asocial). Social learning can be a faster, safer, and in some respects a cheaper way of acquiring representational and/or procedural information compared to asocial learning, sheltering the individual

¹¹ However, note that establishing the social nature of the stimulus may sometimes be difficult. Consider, for example, *behavioural products* as learning stimuli (e.g. scent marks, trails, deposits, material culture templates, discarded artefacts, in sum: all conceivable products of an organism's activity): they clearly have a connection to behaviour, but they're often temporally and/or spatially detached from it. Generally speaking, learning that happens through *products* of behaviour is considered *social* learning (Heyes, 2013b). But there may be unclear cases: for example, does a termite mound count as a social stimulus to a gazelle wandering the savannah? The answer is dubious. Nonetheless, I believe that the distinction between environmental and social stimuli can be maintained by taking into account the degree of ecological relatedness that any specific behavioural product bears to the observer. Very stable and ecologically disengaged products could be considered inert and predictable, closer to environmental than social stimuli.

from the hazards, costs and possible failure of gaining first-hand knowledge. For example, by observing others' behaviour I may learn to dodge dangerous members of my community, as the alpha male in the example; or to avoid poisonous food by indirectly getting to know its adverse effects (Seppänen, Forsman, Mönkkönen, Krams, & Salmi, 2011). Other evolutionary functions speculated for social learning include impacting on populations' evolvability by influencing the supply of behavioural variation (Brown, 2013), and promoting cooperation (Heyes 2013b).

Going back to the two competing definitions of social learning, I think that they are mostly due to alternative disciplinary outlooks on the subject, coming respectively from psychology and cultural evolution theory. Psychology clearly takes the "learning" bit in the locution seriously. Under a psychological perspective, "social learning" primarily denotes a family of cognitive *mechanisms* (more specifically, learning mechanisms), where "mechanism" stands for a fairly coherent cluster of information-processing parts and operations. By contrast, cultural evolution theory takes the mimetic effect of social learning especially seriously, privileging the idea that it can constitute a form of *behavioural inheritance*. Thus the expression "social learning" is often used to characterise the existence of transmission *channels*, carrying information conservatively down generations or circulating it among peers (Sterelny, Smith & Dickison, 1996a).

Recent literature is starting to openly recognise the discrepancy between these two uses of "social learning" and to credit it with importance; through much of the "social learning era" (discussed earlier), however, there was a conflation of "channel" concepts with "mechanism" concepts (Sterelny, 2009). In order to avoid such ambiguity, Hoppitt and Laland (2013), whose book provides the most recent and comprehensive coverage of social learning studies to date, designate the first referent (i.e. channels) simply as "social transmission", reserving the label "social learning" for cases of socially mediated learning (i.e. mechanisms). ¹² I have in principle nothing to object to about this characterisation, which appears to capture the essence of the distinction. In effect, the terms "channel" and "transmission" may be considered conceptually equivalent—and channels are often characterised as "transmission channels". Bearing in mind this substantial equivalence, the "channel" terminology appears nonetheless more convenient here, insofar as it can help to specify distinct "occurrences of transmission", detailing their relationship to social learning mechanisms, and avoiding potential confusion with cases in which the mechanism acts in a non-conservative fashion. The "social transmission" terminology, in fact, seems to obscure the possibility that in some instances transmission is preferentially or disproportionately affected by one particular social learning mechanism rather than another. For example, one conservative channel may be realised by organisms that privilege imitation; or by organisms that are capable only of local enhancement or emulation (or they may be other constraints that favour one social learning mechanism over the other). In these cases, we may speak of an *imitation-channel* or *emulation-channel*. In fact, as we will see, the correct conceptualisation of the RA presupposes the idea of imitation-channels, namely of situations in which when transmission effectively happens, it is disproportionately achieved through imitation. In turn, the fact that it is disproportionately achieved through imitation (rather than through other mechanisms) is supposed to have an effect on the quality of the channel.

Hence, before I go any further, it may be wise to offer a more precise characterisation of what I mean by "channel" and "mechanism" (for the terms

¹² The specific definitions offered by Laland and Hoppitt—which I endorse entirely (though I do not necessarily commit to other distinctions they draw)—are as follows: "Social learning is learning that is facilitated by observation of, or interaction with, another individual or its products. Social transmission occurs when the prior acquisition of a behavioural trait T by one individual A, when expressed either directly in the performance of T or in some other behaviour associated with T, exerts a lasting positive causal influence on the rate at which another individual B, acquires and/or performs T" (Hoppitt & Laland 2013:4).

are widely but rather broadly employed in the literature, cf. Griesemer, Haber, Yamashita, & Gannett, 2005; Griffiths & Gray, 2001).¹³

2.6. Interdisciplinary ambiguity: the channel vs. mechanism distinction

Suppose that Abu has learned about making stone tools from Ina. Abu might have (a) watched Ina's hands and body to get information about how to grip a hammer stone, how high to raise his arm, how hard to strike, etc. Or (b) Abu might have watched the stone that was being worked, noticing the initial shape of the slate, the subsequent concave shape, the sharp edges, the grain of the rock, etc. Or (c) Abu might have watched Ina miming or demonstrating a lithic sequence. Or (d) Abu might have reverse-engineered lithics from a finished artefact made by Ina. Or (e) Abu might have been given a verbal recipe or instruction set by Ina (which, although improbable in this scenario, could also be imagined as written rather than oral).¹⁴ All of these may be described as ways or *channels* through which Ina's knowledge can reach Abu. Such social learning *channels* may or may not correspond to social learning *mechanisms* adopted by Abu.

Channels are defined by realised mimetic outputs, or conservative flow of information. Just as we say that information about eye colour flows through genes to the next generation, we may say that cultural information flows

¹³ Griffiths & Gray (2001) are sceptical of a multi-channelist view of inheritance, but see Griesemer, Haber, Yamashita, & Gannett (2005) for a thorough response and conceptualisation of channels. The discussion between Griesemer, Griffiths and respective co-authors concerns the distinction between (inheritance) channels and inheritance systems rather than channels vs. mechanisms, but the characterization of channels is very similar to the one I offer here.

¹⁴ Also, combinations of (a), (b), (c), (d) and (e) may obviously be imagined.

conservatively through channels. But channels are not metaphors for riverbeds or pipes which continue to exist when the water isn't flowing through. If information stops flowing, channels cease to exist, for channels *are* information transmission, or more exactly *conservative* information transmission (I will imply the "conservative" qualification from now on).¹⁵

Because all that matters to channels is information transmission, channels may be identified as various levels: there may be individual as well as collective channels. Generally speaking, channels are understood to work vertically—i.e. intergenerationally—transporting information from a chronologically upstream point to a downstream one (cf. for example Sterelny, Smith, & Dickison, 1996: 390, describing channels as "routes across generations"), but there is no clear indication as to what boundaries such "information' should have. Ina and Abu's case exemplifies an individual transmission channel, but there may well be collective channels, where a package of cultural information travels through time at the populational level. Imagine, for example, that the transmitted cultural trait is a genuinely cooperative hunting strategy, one for which a certain division of labour is required (Packer & Ruttan, 1988). A cooperative hunting trait clearly can exist only in presence of a thus-and-thus organised cooperative hunting party (Smaldino, 2013). The transmission of such a trait constitutes a collective channel, because the information in question (the hunting technique) depends on the whole group possessing it, and will be generally passed on from group to group (though it may go through bottleneck stages).¹⁶

¹⁶ This depends on the fact that the *idea* of a specific cooperative hunting technique may be stored in an individual brain and from there re-transmitted, like in the putative case of an expert group hunter, who passes on his knowledge to a naïve party. However, such a transmission pattern is unlikely to be

¹⁵ By "conservative transmission" I mean that nodes or relays receiving an information input in the transmission chain deliver the same or similar information as an output. Transmission, by itself, is generally understood as a conservative process, but it is worth highlighting the point because the cultural more than the biological domain may invite ambiguity when information is passed on—or "transmitted"—non–conservatively, i.e. without eliciting the onwards transfer of the *same* information—for example, in communication (Bechtel, 2008a; Morin, 2015; Sperber, 1996).

Even more generally, channels may describe the maintenance of behavioural conformity in a *population* through time (Ramsey, 2013, 2017). Classic cases of "chimpanzee cultures" rely on such usage. "Chimpanzee culture" (or "traditions") denotes cases of behavioural uniformity in geographically isolated populations that lack obvious ecological explanations, and are therefore ascribed to social transmission. The most famous case is perhaps that of different ant-fishing techniques in wild chimpanzee populations at Taï (Côte d'Ivoire) and Gombe (Tanzania). At Taï, sticks are dipped in mounds and termites licked off them directly with the lips; at Gombe, sticks are inserted, termites wiped off with the free hand, and the handful of insects is brought to the mouth. Behavioural similarity identifying the transmission channel is here intended at the level of the whole population. In fact, isolated cases of socially transmitted behaviour would not support the argument about culture: it is intra-group behavioural conformity associated to inter-group dissimilarity that crucially supports the claim (Boesch & Boesch, 1990; Sugiyama, 1985). In this sense, the notion of channel is virtually indistinguishable from that of culture or traditions, clearly indicating that channels can operate at high magnitudes.

Finally, because channels are about *realised* mimetic effects, anything that contributes to the realisation should be considered part of the notion of a channel. If behavioural conformity in a chimpanzee population is thought to depend on demographic factors such as relative proximity between individuals, then demographic factors are part of the channel. As I have mentioned in the first chapter (see objections to accessory claim VII), the study of extra-social

the rule with cultural and behavioural traits. The "cooperative hunting" trait may manage to survive through a few nodes, but it will soon degrade in efficiency and therefore in odds of re-transmission, unless implemented in an actual hunting party (which is necessary for the *expression* of the trait). Typically, behavioural traits' acquisition requires practice, repetition and a modicum of what may be called "embodied cognition"; this seems even truer of collective traits. So while this may be viewed as a case of "silent" transmission, where information goes through one or more nodes in the transmission chain unexpressed (which habitually happens with genetic information), it will be an extreme and unusual circumstance in cultural transmission. learning factors boosting cultural transmission is currently delivering significant results (cf. Andersson et al., 2016; Henrich, 2004; Henrich & McElreath, 2003; Kobayashi, Ohtsuki, & Wakano, 2016; with a modicum of healthy controversy about the exact significance of such factors, cf. Vaesen, 2012; Vaesen, Collard, Cosgrove, & Roebroeks, 2016). Clearly these factors are considered external to social learning *mechanisms*, and part of social learning *channels*. In fact, this stream of work positively highlights the distinction between the two.

In contrast to channels, the notion of *mechanism* is confined to the individual level and does not imply the realisation of mimetic effects. If social learning mechanisms are primarily *learning* mechanisms, they should naturally be interpreted as *cognitive* mechanisms. Unless we are prepared to buy into genuinely collective cognition theories, it seems reasonable to think of social learning mechanisms at the individual level-or at least at the local level, if various learning props and extensions are to be included in the account. Cognitive mechanisms may be said to represent unitary (occurring together and working in concert, presumably with some regularity) clusters of *information* processing tools and vehicles, subserving the production of a specific phenomenon (in our case, learning).¹⁷ The notion of *mechanism* has been widely discussed in philosophy of science, and something like a consensus appears to have emerged from the dedicated literature (see Glennan 2017 and Illari & Williamson, 2012 for the recent "ecumenical" perspective, drawing on previous crucial work from Bechtel & Abrahamsen, 2005; Glennan, 2002; Machamer, Darden, & Craver, 2000). This consensus highlights four main elements of the notion: parts (the different components of the mechanism); operations

¹⁷ "The mechanisms that explain mental phenomena are distinguished from many found in biology as what they must explain is how operations occurring in our head enable us to coordinate our thoughts and actions with the world outside us. Accordingly, they are often characterised as informationprocessing mechanisms because the operation within them produce and change informational vehicles in ways that correspond to the informational content they carry about things and event external to themselves." (Bechtel, 2008b: IX).

(generally intended as processes and interactions among the parts, governed by causality);¹⁸ organisation (ordered interaction of parts and operations); and the *phenomenon* (i.e. what the mechanism produces or is supposed to produce, and should therefore explain; what the mechanism is "for").

For *cognitive* mechanisms, parts may be identified in various ways, including physical elements like the brain and its areas, the body and its extension, but also more elusive entities like motivations, emotions, mental states or representations, or neural gradients. Operations are mostly characterised as information processing (Averbeck & Lee, 2004; Bechtel, 2008a, 2008b): for example, inference may be described as "combining information to generate new information"; memory as "storing information"; understanding as "consuming information". Learning, which is the focal operation here, may be defined as "acquiring information" that can affect behaviour. (But note: the acquisition terminology refers to the *input* bit, not to the acquisition of the behavioural *output*; thus a learning mechanism can in principle be operating in the absence of mimetic output.) Information *processing* may include conservative *transmission* but it is not limited to it; thus mechanisms perform a greater variety of operations with information than channels, which are merely concerned with transmission.

Finally, note that even when mechanisms achieve a mimetic effect, this must be conceived at the *local* or *individual* level: when Abu learns socially from Ina, his behaviour may end up being similar to his master's, but such behavioural similarity is clearly at the *local* level (see next section).

It should be therefore clear that mechanisms and channels are *not* coextensive notions and cannot be reduced to each other. However, they are certainly intimately connected—which may have contributed to the ongoing

¹⁸ It may be worth noticing that, with respect to the underlying notion of causation here required, mechanistic accounts (e.g. Glennan 2002) directed at clarifying scientific explanation have especially tended to endorse the counterfactual or manipulationist view of causation put forward by Woodward (2003).

confusion. While channels and mechanisms certainly need to be distinguished, understanding their connection is even more critical. Although the former may encompass a greater variety of conditions and phenomena than the latter, channels necessarily depend on mechanisms, when operating conservatively, and simply would not exist in the absence of them.¹⁹ In addition, there may be cases in which a social learning channel overlaps entirely with the social learning mechanism effecting the transmission (see for example case (a) in Abu and Ina's example). But various mechanisms may contribute to a single channel, for a trait can be passed on through a combination of cognitive mechanisms (e.g. chimpanzee traditions may be sustained by a mixture of emulation and local enhancement); and transmission channels can include a variety of elements and environmental conditions that do not fit the individual-level dimension of mechanisms (e.g. pine cone abundance, the model's proximity, or the inverse conditions can produce generations of skilled rat pine cone-peelers, or of hopelessly malnourished individuals). In sum, it may be said that mechanisms are necessary but not sufficient for channels; on the contrary, channels are sufficient but not necessary for mechanisms.

In the next section, I examine some problems that stem from failing to properly conceptualise the distinction between channels and mechanisms; then I finally discuss their relationship in the RA.

2.6.1. Confusing channels and mechanisms

In principle, there is nothing wrong with defining social learning or imitation one way or the other—as a channel or as a mechanism—so long as we do so

¹⁹ Note that if behavioural similarity were observed in a population, but depended *entirely* on ecological conditions or other non-social learning factors (e.g. genetic predisposition, developmental biases etc.), we would not speak of a social transmission channel. It is not merely behavioural similarity or mimetic effect that matter to channels, but also their etiology, which must be social.

consistently. Different definitions may simply serve different purposes and fit different explanatory projects. So the problem is not one of the legitimacy of either characterisation over the other, but rather one of ambiguity in failing to distinguish the two clearly. Lack of recognition of the distinction between channel and mechanism accounts of social learning has led to a number of problems (aside from the most general and obvious problem of equivocation, when the meaning of the label "social learning" is changed contextually and inadvertently—or surreptitiously—within the same line of reasoning, cf. Hoppitt & Laland 2013: 3).

There are three serious (and common) consequences of failing to distinguish properly between channels and mechanisms. One concerns the effective realisation of the mimetic effect and may be described, borrowing a useful expression from Morin (2015), as the "flop problem"; the second concerns the level of the mimetic effect; the final one may be defined as the problem of "swapping" definitional criteria for mechanisms and channels. I will discuss them in this order.

The first blunder involves taking the presence of a social learning mechanism as automatic evidence of the presence of a (corresponding) social learning channel. As I said before, social learning mechanisms are generally distinguished on the basis of the informational source in the input condition. However, channels are defined by the value assumed by the output condition— by a *realised* mimetic output. Nothing in the definition of a social learning mechanism automatically implies such realisation (though it may suggest something about how the realisation is obtained). So one social learning mechanism may be present (the organism intakes social information) but may fail to produce a mimetic effect: it seems that the presence of an intermediate stage between input and output conditions, as represented in Figure 2, can highly influence the realisation of the mimetic effect. Motivational factors could intervene and change the organism's course of action away from a mimetic result; time passing and related memory failure may erode the informational

79

content acquired in the input phase. Morin describes this problem as the "flop problem" of social transmission: the fact that we are able to imitate doesn't imply that we always do imitate, by default, or react imitatively to any exposure to social information (in other words, imitation is not, fortunately, "compulsive": Morin 2015). This is also relevant to the problem of interpreting laboratory result about imitation in nonhuman animals: these experiments may at best establish that the experimental subjects *can* imitate (mechanism), not that they *do* imitate (channel) in their ecological settings (Galef, 1990; Galef, 2015).

Secondly, the *level* at which mimetic effects are achieved is important. Again, this has to do with the realisation of the mimetic effect, but at the appropriate level rather than in absolute terms. As I said earlier, even if a social learning mechanism operates conservatively and produces a mimetic effect, this will be at a local level; however, the mimetic effects that define a social learning channel can be, and often are, understood at the *populational* level (cf. Whiten & Ham 1992). "Mimetic effect" is just a general phrase for describing behavioural similarity: but behavioural similarity can happen at various levels, and one level doesn't automatically and linearly translate into another (particularly, the individual level doesn't automatically translate into the populational one: see chapter 4 for an extensive discussion of this topic). After having observed you sticking a twig into a termite mound, extracting it and swiping off the insects in a handful, my actions in feeding on termites may be similar to yours—there will be behavioural similarity at the local level. However, this would be quite a long way away from the level of behavioural similarity involved in chimpanzee traditions, which require the overall population behaviour at time t_1 to be *similar* to the overall behaviour at time t_2 (where t_2 doesn't necessarily represent a generational turnover, but any downstream time slice). The two phenomena are importantly distinct, and although the second depends to some extent on the first, and both can be characterised as "mimetic effects", a number of factors can disrupt the relationship.

80

The RA has often represented a paradigmatic example of this confusion between the potential mimetic effects of imitation at the local level (highlighted by the experimental literature on imitation) and the global level mimetic effects. A cultural channel may be mostly imitative if the mechanism most consistently relied upon is imitation; and certainly we may hypothesise that features of imitation-mechanism can contribute to the quality and success of the imitationchannel. But the overall quality of the imitation-channel is ultimately determined by its *realised* mimetic effect, often intended at the population level, and this will be in turn contingent on many factors other than the presence of imitation-mechanism. First, imitation-mechanism has the potential to generate mimetic effects at the local level, but may not do so (as the flop problem highlights). Secondly, even if local mimetic effects are effectively realised by imitation-mechanism, there will be an array of conditions impacting on transmission at the higher level. Suppose we want to determine the transmission of hand-axe manufacturing in Abu and Ina's group: even if Abu can and does imitate, other factors influence the actual transmission. For example, demographic factors (how many models did Abu have beside Ina, and how many opportunities of interaction with Ina?); environmental factors (how common is the right type of stone? can Abu practice chipping flints often enough?); social hierarchy factors (is Ina a tolerant individual? is he of high rank relative to Abu?), and plausibly many more. If the RA is about explaining cumulative culture, the imitation-channel is presumably intended at the population level. Thus the toolmaking behaviour needs to be reasonably diffused in the population; Abu or Ina need to pass on their knowledge to a sizeable number of individuals. But again: are these other individuals in immediate proximity? How many of them? Organised in what social and demographical structure? And so on.

In conclusion, the RA may give reasons to hypothesise that the type of learning mechanism deployed by Abu (the informational source Abu attends, the imitation-mechanism) has an impact on the channel quality—it may be considered a rightful contributor or a modulating factor. But it is a mistake to confuse the (possible) *local* mimetic effect of imitation-mechanism with the *global* mimetic effect that defines the quality of the channel, even if it is a mostly imitative channel, and assume that one automatically translates into the other.

A final trouble that can derive from failing to distinguish social learning mechanisms and channels properly is "swapping" their definitional criteria. Whereas mimetic outputs legitimately define social learning *channels*, they shouldn't be used to define *mechanisms* (although they can, and sometimes must, be used for *detection* purposes, see below). This problem can be seen as the opposite of the "flop" problem (they are connected): while in that case the presence of a mechanism is used to define a channel, here the presence of a channel is used to define a mechanism.

In fact, this form of ambiguity runs through the whole history of social learning studies: most of the classifications examined in the historical section adopted (more or less covertly) mimetic effects as the lumping definitional criterion, in the same breath as relying on cognitively or psychologically-inspired splitting criteria. This practice produced definitions of social learning and imitation essentially as different ways of "copying". However, if it is social learning *mechanisms* we are talking about, we had better avoid reference to matching output conditions (or at least clarify that matching outputs are a possible but not necessary outcome of learning) and explicitly privilege the input conditions.

This problem is worsened by the fact that we often depend on mimetic effects (the correct definitional criterion for channels) to *detect* and *identify* social learning mechanisms empirically—although these should not be part of their theoretical definition. A particularly clear illustration of this problem comes from the empirical study of nonhuman imitation, where the standard procedure for detection has been the so-called "two-action method".

The two-action method involves exposing experimental subjects to one or more trained conspecifics performing an action, often an instrumental task involving manipulation of some apparatus. The apparatus can be manipulated in two ways (say, twisting the lid of a vase left or right) to obtain a desirable outcome (say, extracting food). The "two actions", or manipulative procedures, will ideally be sufficiently complex (to rule out the possibility that the observer achieves the result individually), and functionally and energetically equivalent in obtaining the outcome (to rule out the possibility that one is preferred for some inherent expediency), but visually dissimilar. In other words, the task is designed in such a way that neither alternative is obviously superior or allows solving the task more efficiently: the only discriminant is visual information stemming from behaviour. Typically, half of the experimental subjects are exposed to a suitably trained demonstrator employing one of the solutions, the other half to a demonstrator employing the alternative. If subjects tend to disproportionately use the alternative that they have observed and produce matching actions, this is taken as evidence of imitation.

Note that these experiments assume a sound definition of imitationmechanism based on the *input* condition. However, such experimental designs run into an empirical difficulty, for even if we assume a definition of imitation based on the input condition, we may have no observable parameter available to assess this condition. As a matter of fact, the output of learning is the only readily observable and measurable parameter in a learning event, and the only way to extract information about the input condition. This is especially true of nonverbal organisms, who cannot express their internal states, or what they have paid attention to, perceived and taken up in the input phase. Thus, matching action *outputs* are used as proxies or *detection* criteria for imitation in nonhuman animals—and legitimately so. The problem is, however, that the distinction between *detection* and *definition* criteria has often been overlooked, resulting in illegitimate "swapping". As a result, the two-action method procedure (which has become widely popular and something of a gold standard for imitation) has contributed to the obscuring of the channel/mechanism distinction in the specific case of imitation, creating the impression that imitation is simply defined as "the mechanisms that produces matching actions".

In conclusion, the confusion between channels and mechanisms can lead to various problems, which are all relevant to the RA, as the next chapters will show. On a very general level, the swapping fallacy can lead to false negative cases, where some social learning is occurring but it is not detectable. These cases include not only cases of non-mimetic social learning like the elephant seal's example, but also cases of potential mimetic learning. These cases fail to realise the consequences due to some intervening impediment (Abu may well have learned how to make stone tools from Ina, but he may never exercise that skill again; he may become crippled or lack the opportunity to work stone). Most importantly, it creates the illusion that imitation-mechanism is *defined* by the conservative and mimetic phenomenon of action matching, and therefore automatically resulting in it (while this is not necessary in imitation-mechanism). This in turn fosters the flop fallacy, which takes "imitation-mechanism" essentially for a synonym of "imitation-channel". The flop fallacy, combined with the assumption that local level mimetic effects translate seamlessly into global level mimetic effects, produces a very typical (and essentially fallacious) RA reading—but this will emerge in time through the following chapters.

It is now time to highlight the relationship between imitation-mechanism and imitation-channel and its relevance in the RA.

2.6.2. The relationship between imitation-mechanism and imitationchannel in the Ratchet Argument

In the previous section, I established that there are two possible ways of interpreting social learning and imitation, and that these two ways should be kept carefully distinct. Both accounts are in principle legitimate, but they perform differently in context, and one might be preferable to the other for certain purposes. It is now time to ask which of the two characterisations we should favour in the context of the RA. Which notion should we plug into this argument: imitation-mechanism or imitation-channel? The answer is mixed, because the argument is effectively *about* both—but I think that imitation-mechanism is primary, because this account of the argument is more informative.

Let us recall first what an imitation-mechanism is. As I have argued in section 2.4, despite various uncertainties and discrepancies in the classification criteria for social learning, a certain convergence may be noted towards the criterion of information source in the input phase of learning. For example, taking the environment as the main source, as opposed to directly taking the behaviour of the conspecific, has provided the main distinction between imitation and local enhancement (and more broadly, between imitation and non-imitative social learning, according to Heyes 1993). Further distinctions about different sources of information stemming from behaviour have produced the imitation vs. emulation classification. The source of information stemming from behaviour can be the outcome of behaviour only (for emulation), or the overall morphological organisation of behaviour, i.e. the specific actions of a behavioural chain (for imitation).

From these characterisations, imitation-mechanism emerges primarily as a learning mechanism sourcing information directly from behaviour, and particularly from the morphological qualities of actions. Although the source of information criterion appears under-determined (in the next section, we will see that it can be divided further, and that a wealth of informational sources can stem from behaviour), I think this should be considered the standard and basic characterisation of imitation-mechanism.

As for imitation-channel, instead, I proposed that this notion may identify cases of *realised* transmission, obtained primarily through imitation-mechanism. I then recognised a number of problems that may prevent imitation-mechanism from resulting in an actual imitation-channel. These may be summarised as (a)

85

local features of the learning episode, such as motivational or memory failures, resulting in local "flop problems" (i.e. imitation-mechanism may be present, but no local channel is formed); and (b) global factors preventing the local channel from translating into a population-level, more properly "cultural" channel. So there may be three elements in play: imitation-mechanism (always local, by definition); local imitation-channel (i.e. an imitation mechanism that has solved the flop problem); and global imitation-channel.

Now, if imitation-mechanism never had mimetic effects at the local level (which would be strange), it wouldn't make sense to consider it relevant to the RA, which is about cultural transmission, and therefore about channels. So in general, for the purpose of discussing the RA, we need to assume that imitationmechanism solves at least the flop problem, in the cases under RA examination, resulting in local imitation-channels. This seems to push us back to a definition of "imitation" as copying—perhaps slightly more qualified as: copying resulting from attendance of the informational source of actions morphology. This is not a small qualification. As formerly noted, it is not unreasonable to expect that attending to behaviour directly may change the quality of the local transmission channels in various ways. Adopting a mechanism-centred definition of imitation affords us a way of evaluating the frequency of the potential occurrence of "copying", their reasons for failure, and their promoting conditions. It offers a much more sophisticated and qualified view of the local imitation-channel.

In the rest of this work, and for the sake of the RA discussion, I am therefore going to assume that we are dealing with cases in which imitationmechanism has effectively solved the flop problem, i.e. it has overcome a possible obstacle in the intermediate phase of learning and has resulted in locally matching behaviour. I will therefore use the expression "standard imitation" (or simply "imitation") to indicate both imitation-mechanism and *local* imitation-channels.

I am also going to make a further crucial assumption: namely, that the source of information selected by any particular mechanism in the input phase

(e.g. morphology of behaviour, outcome of behaviour, etc.) has a direct bearing on the output phase, i.e. on the resulting "style" of copying or local channel. That is, if "imitation" is defined as sourcing information from the morphology of behaviour, it will result in preferentially copying the morphology of behaviour; if "emulation" is defined as attending to the outcome, it will result in copying the outcome ... and so on. This is not an entirely accurate assumption; the reason I adopt it is that this is the standard approach in the literature.

Notice finally that assuming that the RA deals with the subset of cases in which imitation-mechanism has turned into a local channel doesn't subtract from the interest of the line of reasoning. The quality of the *global* imitationchannel will still depend on a host of factors; the job of the RA will be to spell out and evaluate the specific effects of imitation (i.e. mechanism and *local* imitation-channel) on the peculiar form of culture (i.e. the global channel) deriving from the preferential adoption of this social learning mechanism.

This lengthy discussion about channels and mechanisms has been intended to point out that this task has often been overlooked due to the failure to clearly differentiate between the two notions, with the consequence of disregarding the levels and actuality of mimetic effects and/or committing the "swapping" fallacy.

2.7. Intra-disciplinary ambiguity about imitation

With the current chapter coming to a conclusion, the initial ambiguity of the notion of imitation in the RA has been reduced. As formerly noted, imitation as *all copying* is not a good candidate definition for imitation in the RA; neither is the unqualified notion of *imitation-channel* (which must instead be qualified as indicated). The account of imitation that we want to use in the RA is thus reduced to a specific case of social learning *mechanism* discerned from others by

appeal to input-based criteria. As I have remarked, however, the criterion concerning the source of information needs to be detailed further. So far, I have distinguished only the possibility of behavioural vs. environmental sources of information, and within the behavioural source the one of outcome vs. action morphology. This gave us a standard definition of "imitation" as attending to behaviour directly and specifically sourcing information from action morphology—the imitation vs. emulation distinction; but it also suggested that behaviour may be a yet richer source of information.

This last section isolates four extra notions of "imitation" (which again may go under slightly different names in the plethora of currently available classifications) whose conceptual profile can be further distinguished. All these mechanisms fit the definition of *attending to behaviour directly*, as opposed to attending to information from the environment, and have been in at least some cases identified under the label "imitation" (with perhaps further specifications). In addition, some of these mechanisms *also* fit the standard definition of "imitation" as attending more specifically to the *morphology* of actions; but others propose informational sources importantly *different* from behavioural morphology. The mechanisms are the following:

- a) "True" imitation, or "mentalised" imitation
- b) "Simple" imitation, or mimicry
- c) Over-imitation
- d) Program-level imitation

As I have remarked in the previous sections, classifications of social learning appeal to confusingly heterogeneous criteria. Thus these four mechanisms could be grouped in various manners, depending on which criteria we adopt to draw the distinction. If we adopt the criterion concerning the source of information as primary, as I suggested, we obtain that over-imitation and mimicry are subsets of the standard definition of "imitation". These two mechanisms imply attending to morphology for the most part, but respond also to further criteria (level of intentional control and understanding of causal relevance of information).

Instead, always under the source of information criterion, program-level imitation and mentalised imitation must occupy different slots. Roughly, the first might be characterised as sourcing *causal* or *procedural* information from behaviour; while the second might be characterised as sourcing *intentional* information.

This is the main classification guide that I will propose for usage in the RA. However, this classification obscures other important features of these mechanisms. So before plugging them back into the RA, I take a final modest detour and discuss them in pairs, according to different but perhaps more informative criteria found in the literature.

I start with discussing "true" or mentalised imitation, and "simple" imitation or mimicry. These two mechanisms are in many ways polar opposites. With respect to the criteria of classification discussed so far, which were focused on the *input* phase and the intake of information, these two mechanisms are distinguished on the basis of the cognitive machinery that is thought to operate in the *output* phase (warning: *not* on the mimetic or non-mimetic *value* taken by the output phase).

All realised instances of imitation/imitative phenomena face the challenge of explaining how the information acquired in the input phase is used to guide action in the output phase. This problem goes under the name of the "correspondence problem" in the contemporary literature (Brass & Heyes, 2005). The difficulty lies in explaining how the *visual* representation of the model's actions can be subsequently turned into a *motor* representation for directing the observer's behaviour (visual and motor representations are traditionally depicted as distinct modalities of mental phenomena). Solving the correspondence problem and explaining the inter-modal flow of representations is thus necessary for explaining the performance of any imitative phenomena. Mentalised imitation and mimicry differ principally (though not only) in

89

postulating, respectively, a very complex and a very simple way of solving the correspondence problem.

2.7.1. True imitation, or mentalised imitation

The expression "true imitation" is generally attributed to Tomasello (Tomasello et al., 1993). Sometimes the distinctive criterion for true imitation is merely indicated in the *novelty* of the model's action. The idea is that if an action was previously in the repertoire of the observer, the model's performance may simply elicit or make more probable the observer's matching performance, making it harder to maintain that the observer has truly "learned" (acquired) something and "truly imitated". According to some, such a phenomenon is better understood as contagion, priming or social facilitation (Byrne & Tomasello, 1995; Byrne, 1999; Zentall, 2001). Tomasello's initial characterisation of true imitation included reference to the novelty criterion. However, I think that its most important feature is, rather, the requirement of theory of mind abilities, as emerging from his influential 1993 paper: "True imitative learning in our definition involves the infant's reproducing the adult's actual behavioural strategies in their appropriate functional contexts, which implies an understanding of the intentional state underlying the behaviour" (Tomasello et al., 1993: 497, my Italics).²⁰

This cognitively rich view of imitation is perhaps a legacy of early investigations on the topic. Early literature viewed the correspondence problem as exceptionally challenging, and presupposed that in order to solve it, some

²⁰ It is fair to say that Tomasello's imitation has always been overly "mentalised"; in fact, it becomes even more so in later work, where imitation is suggested to depend on "shared intentionality" (Tomasello & Carpenter, 2005; Tomasello, Carpenter, Call, Behne, & Moll, 2005). But remember that the RA discussed here represents a broad form of orthodoxy used in the field of cultural evolution, rather than the original argument formulated by Tomasello, so his definition is not binding.

capacity to "translate" mental representations, or to "see the world from the others' perspective" was needed (Heyes 1996; Whiten & Ham 1992: 271; Byrne 1999:63). In other words, the translation between visual and motor representational modalities was thought to depend on the ability of the observer to form abstract, flexible, and high level mental representation of actions (i.e. meta-representations), which would allow her to take "a relatively low-level visual representation of an observed action, recoding it 'up' into an abstract representation—a symbolic code, program, perspective or intention—and then recoding it 'down' into a motor program" (Heyes 2013b: 315).²¹

Should mentalised imitation be considered relevant to the RA? I am tempted to say no, although this "imitation" is in effect the sort of mechanism that Tomasello first envisaged (and over time enriched). If the one described above were the only way of solving the correspondence problem, we should expect to find "imitation" only in species endowed with fully-fledged theory of mind abilities—in short, only in humans and very likely only in late humans. This expectation has been spectacularly disappointed. In the post-RA period, imitative abilities have been demonstrated in a number of taxa, including primates, rodents, birds, reptiles and even some invertebrates (Akins & Zentall, 1996; Buttelmann, Carpenter, Call, & Tomasello, 2007; Fawcett, Skinner, & Goldsmith, 2002; Fiorito & Scotto, 1992; Heyes & Saggerson, 2002; Heyes, Dawson, & Nokes, 1992; Kis, Huber, & Wilkinson, 2014; Voelkl & Huber, 2000). This evidence suggests that the machinery required for solving the correspondence problem, and therefore for performing imitation, is probably relatively simple, cognitively unsophisticated, and phylogenetically widespread. Notice that this doesn't deny that high level mechanisms such as intention reading or the like are *ever* involved in imitation: it simply proves that they are

²¹ Other candidate solutions speculated for the correspondence problem include "symbolic coding", "program extraction", "perspective taking" and "intention reading", cf. Heyes 2013a.

not required.²² However, I will include the mentalised picture of imitation in the Abacus scheme for the sake of completeness. This interpretation of imitation, it goes without saying, makes the RA an anthropocentric line of reasoning, and (even worse) a late-anthropocentric one.

2.7.2. Simple imitation, or mimicry

Let us now look at simple imitation. Simple imitation is sometimes also called "mimicry", or "the chameleon effect" (Cook, Bird, Lünser, Huck, & Heyes, 2012, Heyes 2013b, Chartrand and Bargh 1999, Dijksterhuis 2005). It generally describes the tendency to mirror others' actions, postures, manners, facial expressions, often intended as low-level motor schemes that are already present in the observer's repertoire. Examples of simple imitation are the tendency to walk at a slower pace in the presence of the elderly (Kawakami, Dovidio, & Dijksterhuis, 2003), to scratch one's nose or dangle one's feet when one's interlocutor is doing so (Chartrand & Bargh, 1999), as well as neonatal facial imitation, emerging as early as in the first hours after birth (Meltzoff & Moore, 1977, 1983, 1997). Mimicry (I will prefer this term purely for brevity reasons) may be considered antipodal to the mentalised brand because it is generally depicted as depending on simple, largely unconscious cognitive processes and operations, most often some automatic cross-modal matching device (matching visual to motor representation). Mirror neurons (whether "built by experience" or innately hardwired, see below) represent perhaps the most popular and widely endorsed hypothesis in this sense, identifying a class of cerebral cells or areas that respond concomitantly to the performance and observation of

²² For the definition of "imitation", in turn, it seems to imply that what happens in the output phase is less relevant to identify the mechanism—when behavioural morphology is used as a source of information to produce matching behaviour, whether through a simple or complex cross-modal mechanism, we are willing to speak of "imitation".

particular motor acts (Gallese, Gernsbacher, Heyes, Hickok, & Iacoboni, 2011; Iacoboni, 2009; Ramachandran, 2000).

Should mimicry be considered relevant to the RA? I think it should. In essence, the existence of something like mimicry corroborates the idea that the cognitive mechanisms recruited to solve the correspondence problem are active at a fairly low level, requiring no complex meta-representational capacities or even intentional control. As a matter of fact, we still lack a satisfactory explanation of the fine-grained neural and cognitive machinery at work in the cross-modal context; the mirror neurons literature is rich and expanding, but at present vexed with controversies. Nonetheless, my hunch is that mimicry could capture the existence of a simple cognitive mechanism of cross-modal matching, rather than isolating some special category of social learning. Roughly, mimicry or simple imitation could probably be considered the rudimentary machinery underpinning all the imitative mechanisms described so far.

Two qualifications and a final consideration must be made.

First caveat. This take on mimicry does not imply that whatever mechanism makes mimicry possible constitutes must be "innate" or "inbuilt" (let alone an adaptation): the origin question is a separate one (Cook, Bird, Catmur, Press, & Heyes, 2014; Heyes, 2010a,b). With respect to the innate vs. learned controversy that appears to be the current crux of interest in cross-modal matching mechanisms (Gallese, Gernsbacher, Heyes, Hickok, & Iacoboni, 2011; Heyes, 2010a,b; 2013a; Hickok & Hauser, 2010), I wish to take no side. I think in fact that the available evidence is inconclusive; we simply don't know enough yet.

Second caveat. By postulating that mimicry describes the basic mechanism of cross-modal matching lying beneath all imitative phenomena, I do not mean that it is equivalent, completely overlapping, or coextensive with the latter phenomena (much as it doesn't mean that all imitation happens exclusively through this mechanism). Instances of imitation such as the reproduction of complex feeding, manufacturing, and manipulative tasks

93
described throughout this chapter will almost certainly require *other* cognitive machinery and explanation. Conscious imitation of complex tasks will have higher demands on executive processes, working memory, and in general computational capacities. However, as in Heyes' words, "the challenge posed by the correspondence problem does not vary with sequence novelty", or at any rate with the complexity of the copied task (Heyes 2013b: 315). In other words, the correspondence problem can be considered "once solved, always solved": the cognitive difficulty of complex imitation may simply lie elsewhere.

Finally, there is another interesting aspect of mimicry that is worth mentioning and may play a role in the RA. Mimicry has been connected to a "social glue" effect (Chartrand & Bargh, 1999; Chartrand & van Baaren, 2009; Heyes, 2013b; van Baaren, Janssen, Chartrand, & Dijksterhuis, 2009). A small cottage industry of psychological studies has flourished around the idea that "imitation leads to liking" (Dijksterhuis, 2005: 209)—or at least this effect has been clearly documented in humans.²³ Something similar to such an effect seems to hold also for animals, where it has been described as "contagion", "social facilitation" and "priming", occurring for example in herding and flocking, as well as in mobbing and appetitive contexts (Byrne & Russon, 1998; Zajonc, 1965; Zentall, 1996; 2006). Although this possibility has rarely been taken into account (but Heyes 2013b is an exception), I suggest that this prosocial secondary effect of mimicry may play some explanatory role in the RA, especially by helping to solve the "flop" problem about lack of motivation to imitate.

²³ Perhaps the relationship is better characterised as a virtuous circle, where imitation leads to liking and liking leads to more imitation (Heyes 2013b). In effect, early studies on the topic highlighted only a correlation between the two, without clarifying the direction of causality; later ones investigated the phenomena in detail, confirming a virtuous circle dynamic, but also a clear causal role for imitation in promoting prosocial feelings (Chartrand & Bargh, 1999; Lakin & Chartrand, 2003).

2.7.3. Over-imitation

Let's finally examine over-imitation and program-level imitation. These two mechanisms also, I think, are helpfully discussed in contrast to one another. Here we need to shift our attention back to the input phase (specifically to what kind of information the observer is sensitive to, and becomes part of the mental representation that guides output behaviour), and to the motivational factors of learning.

Over-imitation, also known as "dumb", irrational, or mindless, is defined as the tendency to copy the morphology of observed action to an extremely fine-grained level of detail, but ignoring the distinction between actions that are causally relevant to the task and those that are not (Heyes, 2013b; McGuigan, Makinson, & Whiten, 2011; Whiten, McGuigan, Marshall-Pescini, & Hopper, 2009). Some of the best documented cases of over-imitation involve human children observing a demonstrator opening a box-device to retrieve a reward inside (Whiten et al., 2009). In such experiments, opening the box typically requires a complex series of actions such as unpinning bolts, lifting doors, pulling locks. Children are far more likely to copy *all* of the demonstrator's action, even the patently irrelevant one, such as for example tapping on a bolt before disabling it. The same result was observed even when the box was made of transparent material, which rendered the causal pointlessness of some actions even starker and more surprising.

Over-imitative phenomena have roughly two possible explanations (Lyons, Young, & Keil, 2007). One is that the observer does not *understand* the difference between causally relevant and casually irrelevant actions, so intakes and copies pedantically whatever she observes (Lyons, Damrosch, Lin, Macris, & Keil, 2011). The other is that she understands the difference between causally relevant and irrelevant actions, but for some reason ignores it when it comes to copying the actions. In other words, the observer *chooses* to uneconomically over-imitate (Nielsen, 2006). The second interpretation appeals to motivational

factors. But if we take motivational factors into account, one would expect the observer to get to the goal as fast as possible (e.g. opening the box and retrieving the prize), ignoring fiddling with the irrelevant actions. Over-imitation challenges this prediction, and suggests that children may prioritise motivations other than that of retrieving the prize.

Both possible explanations of over-imitation reveal another interesting assumption about the sources of information stemming from behaviour. By contrasting standard imitation with emulation, only one possible distinction about the sources of information in behaviour was drawn: the observer may preferentially attend to the outcome or to the morphology of actions. With overimitation, a new informational layer of causal understanding emerges (Horner & Whiten, 2005; McGuigan et al., 2011). This suggests that the informational source represented by behaviour is not a mere assemblage of perceptual inputs about spatial and temporal organisation of actions, but can be read at a higher level—at the *causal* or *procedural* level (and perhaps at other levels, cf. Call & Carpenter, 2002). In fact, I think that other sources of behavioural information may be recognised and identified, along with procedural and morphological, temporal, intentional, semantic and possibly transitive sources of information in behaviour.

Should over-imitation count among the mechanisms that are relevant to the RA? I think the answer is yes. Over-imitation fits the general definition of using behavioural morphology as the source of information to potentially produce matching behaviour. Often, indeed, over-imitation is taken to represent an extremely pedantic form of morphological imitation (something like an extraaccurate imitation). In addition, over-imitation is disproportionately observed in humans (Clay & Tennie, 2017; Nielsen, 2006; Nielsen, Subiaul, Galef, Zentall, & Whiten, 2012; Nielsen & Tomaselli, 2010) while occurring in other animals (especially apes) only under certain conditions (Horner & Whiten, 2005; Whiten et al., 2009). The tendency to pedantically prefer the morphological source over the apparently more rational choice of procedural information will require explanation; but such explanation is likely to contribute, I think, to the understanding of the RA, perhaps in connection to the prosocial role identified before for mimicry.

2.7.4. Program-level imitation

The last brand of imitation I wish to discuss is so-called program-level imitation. This definition is from Byrne (Byrne, 2003; 1999; Byrne & Russon, 1998), and, surprisingly, it has been largely and I think unduly neglected (although sequence or production imitation, two closely connected notions, are recently gaining traction (Hoppitt & Laland, 2013; Nguyen, Klein, & Zentall, 2005). Byrne's definition stems from the observation of imitative behaviour in great apes, particular mountain gorillas (Gorilla beringei beringi) living on the high slopes of volcanic ranges in Rwanda, Zaire and Uganda. Gorillas occupying this subalpine habitat must rely on a diet of leaves and pithy stems of herbaceous plants, particularly stinging nettles (Laportea alatipes) and bedstraw (Galium ruwenzoriense), which are covered in tiny hooks. Gorillas have developed rather complex processing techniques for eating these plants while avoiding the nasty defence mechanisms. They strip handfuls of leaf blades from the plant, twisting the stems off, removing possible debris from half-open hand, and folding the handful over the thumb in such a way that the nastier plant edges are trapped inside a "leaf-sandwich" (this is the technique for nettles specifically, cf. Byrne & Russon 1998). Although no experiment has been carried out on this critically endangered species, hundreds of hours of field observation provide good reasons to think that such behaviour is learned socially, mainly through contact with the mother and the dominant silverback (Byrne & Russon, 1998).

The "program-level" label has been proposed because while young gorilla learners exhibit *no* matching in the low-level components of the behavioural

technique (such as, for example, hand preference, manual style, and other idiosyncratic aspects of movements), a high degree of matching is observed in the high-level components of the technique (i.e. in its causal, procedural nodes, such as producing a whorl of leaves, a whorl of stemless leaves, a parcel of folded nettles in this relative order). Byrne and Russon note that given the large amount of possible combinations of such high-level elements of the technique which could be achieved by trial-and-error learning and equally succeed in feeding the individual, it is highly implausible that all animals would end up with the same sequence of actions if it were not for some imitative phenomenon. In the authors' words, "program-level imitation may be defined as copying the structural organisation of a complex process (including the sequences of stages, subroutine structure, and bimanual coordination) by observation of the behaviour of another individual, while furnishing the exact details of actions by individual learning" (Byrne & Russon 1998: 676).

The functioning of program-level imitation may be contrasted to that of over-imitation: in over-imitation, action morphology is pedantically copied down to the level of causally irrelevant details; in program-level, action morphology is ignored and causal or procedural information is the target of attention.

I think that on a more general level we should focus on the fact that the type of informational source selected by program-level imitators is profoundly different from the one usually postulated for standard imitation, i.e. action morphology. Causal or procedural information may be contrasted with morphological information in clearly requiring a higher level of cognitive processing; as in the case of mentalised imitation, some ability to extract complex representations from the observation of behaviour re-enters the scene. However, the processing of causal information: it certainly doesn't have to postulate theory of mind abilities. Accordingly, program-level imitation has been tested in primates and birds (to my knowledge, Byrne, 2002; Nguyen et al., 2005); although not in other taxa, due to the fact that designing sequential tasks

with appropriate controlled conditions that are at the same time cognitively manageable by other species has proven difficult.

In particular, the processing of procedural information may even dispense with the hypothesis of the complex "up-to-the-abstract and down-to-the-motor" route of inter-modal matching required by mentalised imitation for solving the correspondence problem; if the mirror neuron research is correct, it suggests evidence for a simpler neuronal mechanism of "action understanding" and matching action production (Butterfill & Sinigaglia, 2014; Caggiano et al., 2012; Kohler, 2002; Lepage & Théoret, 2007; Rizzolatti, Fogassi, & Gallese, 2001; Umiltà et al., 2008).

Should program-level imitation count among the mechanisms that are relevant to the RA? This matter is complicated. On one hand, I think that the importance of procedural information in explaining cumulative culture has too long been ignored, and is likely to be fundamental (as indicated by Byrne and Russon 1998). On the other hand, it is unclear whether the available conceptualisation of program-level or sequence imitation is likely to be of any help—and such conceptual difficulties might explain why program-level imitation has attracted remarkably little attention, despite clearly being an interesting element of the picture. Byrne and Russon in fact suggest that program-level imitation is the fundamental imitative mechanism for primates and humans alike, and downgrade all other imitative phenomena (roughly speaking, imitation at the level of actions rather than program) to some sort of "behavioural priming" or social facilitation (Byrne & Russon 1998: 683). This conclusion doesn't sit well with the evidence for over-imitation and seems in general to underestimate the available evidence about the correspondence problem. Nevertheless, I think that we should strive to fit program-level imitation, or some equivalent process, into the RA picture, perhaps by giving more careful consideration to a connection between low-level perceptual information and meaningful, high-level procedural information.

2.8. Conclusion

In this chapter, I have discussed the historical evolution of the notion of imitation and highlighted a number of conceptual difficulties associated with it. While the main aim of the chapter was to identify and isolate as sharply as possible the different notions of imitation that have been considered relevant to the RA, I definitely haven't resisted the temptation to editorialise. The result of the analysis is summarised in figure 11, which illustrates some of the conceivable conceptual paths one can walk to discriminate imitation from the greater set of social learning mechanisms (indeed, it depicts the paths that I have chosen to walk in this chapter).

The result of this analysis is the distinguishing of six different meanings of "imitation" that will be plugged back into the RA Abacus scheme. All of them have been mentioned or gestured at in connection to RA inferences on at least some occasions in the literature (Arbib, 2011; C. M. Heyes, 1995; Heyes, 2013b; Shipton, 2010) although of course some of them will be more commonly referenced. What I call "standard imitation", i.e. attending primarily the morphology of actions and resulting in so-called action matching, is by any means the most prolific account (known also as process-copying, copying the strategy as opposed to the outcome only, copying the topography of action, and various others expressions). The others have more modest roles, but are in some cases, as we will see in the next chapters, overtly taken as the relevant RA terms. Perhaps the most neglected account is Byrne's program-level imitation. This neglect is undeserved: I will return to discuss a possible role for program-level imitation in the final chapter.



Figure 11. Different categorisation paths examined in this chapter and the resulting accounts of imitation in the RA.

CHAPTER 3. What is Cumulative Culture?

3.1. Introduction

The second term of the RA is cumulative culture. Again, as for the other terms, it is important to clarify what we mean exactly by this term. Under certain accounts of cumulative culture, the RA borders on triviality and irrelevance, while under other accounts it appears like a highly contested hypothesis. Thus this chapter, like the previous one on imitation, is largely devoted to teasing out different notions of cumulative culture associated with the RA, while clarifying conceptual hurdles along the way. It may be important to emphasise preliminarily that the goal of the chapter is not to evaluate the absolute merits of different accounts of cumulative culture (e.g., empirical reach, internal coherence, comparative significance), nor to defend a particular account of cumulative culture. The main objective is instead to identify different notions of cumulative culture which, when incorporated in the RA as explanatory targets, have encouraged different interpretations of the argument's logic. Hence I will give particular attention to those features of the definitions of cumulative culture, often implicit and overlooked, that have significant bearing on RA's readings.

The chapter opens with the discussion of simple or plain culture. The notion of *culture* is foundational: it logically precedes the one of cumulative culture, insofar as the latter is a specific brand or subset of the former. It

therefore seems sensible to get a steady grip on what we mean by "culture" beforehand. I will take into account only evolutionary informed definitions of culture; here, a solid consensus exists around a broad and minimal definition (roughly speaking: culture is anything that gets repeatedly socially transmitted, see Laland & Galef, 2009). As a result of this broad definition, the term "culture" applies to material objects (artefacts, tools, ornaments, and more broadly any engineered portion of material world—a path or a stone dam, for example), to behaviour (skills, manufacturing, communicative gestures, ritual or symbolical practices such as dance and tattooing), as well as to insubstantial entities like information, ideas, designs, recipes, norms, tales (including language-encoded ones). Most importantly, culture can legitimately identify entities at the local/individual as well as at the global/population level, being in this sense a level-neutral notion; and while an informal distinction among "cultural domains" is virtually omnipresent (e.g., social domain, technological, foraging, symbolic domain, etc.), the term "culture" is regularly used to denote all these domains comprehensively and indistinctly.

I then briefly discuss historical, pre-RA notions of cumulative culture. In contrast to the case of *imitation*, historical ambiguity is limited and can be essentially set aside: the idea of cumulative culture is largely coeval with the RA. It is however interesting to note that broad-brush incarnations of cumulative culture had been around before, and that they designated with surprising accuracy the same kind of phenomenon Tomasello intended to explain with his appeal to imitation and fidelity. What distinguishes historical from post-RA definitions thus is not their referent, but rather their formulation. Historical accounts are generally cashed out in terms of anthropocentric examples and suggestive analogies rather than objective criteria; contemporary definitions of cumulating definitions in an empirically- and comparatively-friendly manner. Operational definitions require moving away from *examples* (essentially human-based) towards *unified*, *objective criteria* for identifying cumulative culture empirically.

The need for operational (empirically tractable, criteria-based) definitions stems from the fact that the RA transferred the notion of cumulative culture into a more decidedly evolutionary framework, embedded in the developing conceptual landscape of cultural evolution. How well modern definitions of "cumulative culture" may have met this objective will not be a matter of serious concern here. Instead, I will focus on identifying the accounts of cumulative culture that emerged from this process of operationalisation and attempted unification.

The discussion of modern criteria for cumulative culture can be, at a very general level, organised around a simple chronological dichotomy. In recent years (circa 2009 onwards), one single criterion appears to have taken up the lion's share of attention, and may be said to represent the current consensus on cumulative culture. I will label this recent criterion *recapitulation*. However, before the sweeping success of recapitulation, the conceptual landscape of cumulative culture was more fragmented: it can be roughly divided into two alternative ways of thinking about cumulative culture, surrounded by a nebula of variously associated criteria (3.4). In the absence of better labels, I identify the two principal and competing pre-recapitulation accounts of cumulative culture as *accumulation of traits* (or "Jablonka's definition"). These accounts both interpret cumulative culture as culture that, by accumulating modifications over time, becomes more complex and more efficient; but due to the ambiguity about levels, they end up identifying quite different processes.

Most of the conceptual work in the chapter is done in contrasting Tomasello's and Jablonka's accounts of cumulative culture and their fundamental components: accumulation, complexity and efficiency. (I do not discuss in detail the other minor pre-recapitulation criteria, for, as will become clear, these represent diagnostic signs or corroborative evidence of cumulative culture, rather than identifying criteria.) The divergences between these two accounts are important because they implicitly introduce mismatching meanings of *complexity*, *accumulation* and *efficiency* into the cumulative culture debate. Tomasello's account represents the mainstream view on cumulative culture; but even though most authors explicitly embrace Tomasello's interpretation, many of them fall back or tacitly switch between different usages of the aforementioned notions, leading to significant equivocations in the evaluation of the RA.

The discussion of complexity and efficiency will lead me to advocate for a sharp separation between "general" cumulative culture (including traits from both the technological and non-technological domains) and cumulative *technology*—which may be also seen as two alternative accounts of cumulative culture plugged in the RA. As it happens, the RA can afford us a rather interesting explanation of cumulative technology, while it is probably only shallow if its target is general cumulative culture.

What about recapitulation, finally? While representing an improvement in the debate (for it zeroes in on single traits, avoiding level ambiguity), I will argue that recapitulation is no more than a cleverly compact rephrasing of Tomasello's definition. This has two consequences: on one hand, it straightforwardly imports many of the conceptual hurdles of pre-recapitulation accounts-so the discussion of the latter is both necessary and valid for the refurbished criterion as well. Secondly, in terms of relevance to this work, recapitulation occupies essentially the same position in the discussion of the RA; thus, as accumulation of modifications, it should be understood differentially for technological and non-technological domains. In essence, recapitulation recommends labelling as "cumulative" only those cultural traits that go beyond individual inventiveness that are too *complex* for one individual to invent. The criterion hence capitalises on an epistemic notion of complexity, which conveniently allows "recapitulating" the social history of modifications in a trait at the level of individual cognition. But recapitulation is only deceptively simple. The criterion is designed to imply that a trait obtains its cognitively complex form through the concerted efforts of different individuals and/or a history of preserved modifications, which in most

cases remains the necessary evidence for a trait to be deemed a genuine instance of cumulative culture, especially for field observations. Thus recapitulation may appear at first sight like a unified and operational golden standard for the detection of cumulative culture (which incidentally, doesn't support evidence of cumulative culture outside humans); but in reality it needs to be "unpacked" into its ancestral version with all the related trimmings, in order to be effectively applied—which causes its empirical applicability to plummet.

I think this result is scarcely surprising. In the process of ditching its anthropocentric bias and looking for unified and empirical criteria, the notion of cumulative culture inevitably comes to harbour a fundamental tension. Cumulative culture is at its core a *human* phenomenon: evidence in nonhuman animals is at best circumstantial, extremely scarce, and highly controversial;²⁴ evidence in humans is self-explanatory and archetypal. The search for a unifying, operational criterion to define cumulative culture must expect partially unsatisfactory results on either side. Consider, for example, the following banner cases for cumulative culture: the Eagle lunar module of Apollo 11 spaceflight (Dean et al., 2014); a 500,000-year-old hafted spear (whose stone tips were found at Kathu Pan 1, South Africa, cf. Wilkins, Schoville, Brown, & Chazan, 2012; Haidle & Conard, 2011); "stepped" probing tools manufactured by New Caledonian crows out of Pandanus leaves (Hunt & Gray, 2003), pictured below.

²⁴ In fact, the major comparative review of cumulative culture available to date concludes rather categorically: "Presently there is no evidence that any species, except humans, have cumulative culture" (Dean & Laland 2014: 16).



Figure 12. Three banner cases of cumulative culture (1) the Eagle lander of Apollo 11; (2) stone spear tips found at Kathu Pan 1, South Africa; (3) Pandanus leaf tools fashioned by New Caledonian crows

What criterion can possibly hold these three things together and describe them in a unitary and satisfactory way? If it can be found, it must be a very general one, perhaps apt to fully describe the simplest cases, but capturing only some basic features of the most complex ones.

Modern criteria for cumulative culture aspire to unity but are distributed across a spectrum that goes from minimal to rich. Some criteria, like diversification, pre-planning, or composition (see section 3.4), are crafted to provide a rich and fully-fledged (human-like, one would be tempted to say) description of cumulative culture; others (like recapitulation) appear to be inspired by the minimalistic project of identifying empirically the essential features of a cumulative culture trait.²⁵ In some respects, these criteria are designed to serve fundamentally different *questions* about cumulative culture. The first type assumes that humans have cumulative culture in prototypical form, and asks how such cumulative culture is sustained and how it evolves,

²⁵ A similar tension has also been remarked in the foundational notion of culture with the distinction between "thick" and thin" (and "medium viscosity") definitions of culture (Wimsatt & Griesemer, 2007).

rather than how it originates; what are its specific dynamics; what it looks like in humans. It will often include typical but unnecessary conditions in the account, and focus on process description rather than cumulative culture detection. The second type is interested in dawning forms of cumulative culture, and looks painstakingly at the bootstrapping conditions that mark the difference between simple culture and cumulative culture; it might overlook or downplay fundamental features associated with human cumulative culture such as gradualism and the advent of the symbolic culture.

There is no fundamental contradiction among these different criteria, i.e. they don't produce incompatible definitions of "cumulative culture" (although they might produce *incomplete* definitions); thus I will not return to this topic explicitly in this chapter; the issue simply needs to be mentioned as a preliminary to put the remainder of the discussion into correct perspective. It is indeed unreasonable to expect a single criterion to satisfactorily capture dawning forms of cumulative culture as well as truly complex, fully-fledged variants. As Heyes puts it, in fact, both "the grist and the mills" of culture are subjected to evolution (Heyes, 2012b; for similar ideas about "dynamic" views of "evolving evolution", see also Calcott & Sterelny, 2011): cultural products evolve (the grist), but so do the very processes of cultural evolution (the mills), altering their rules and dynamics over time. The issue will come down in most cases to evaluating on a case by case basis what criterion is more appropriate for the situation at hand. Keeping in mind this double-pronged research focus on cumulative culture, nonetheless, should help to put into meaningful perspective the potentially overwhelming variety of criteria associated with it.

As far as the RA is concerned, the different notions of cumulative culture plugged into it have been at times of the rich kind and at other times of the minimal kind; although perhaps it is fair to say that the original version of the argument, and certainly one of its most interesting (though narrow) interpretations, assumes a fairly rich, gradualistic and technologically-biased definition of cumulative culture (see chapter 5).

3.2. Plain culture: one evolutionary perspective, many referents

In this section I present the broad consensus reached by evolutionary studies on the definition of "plain" (i.e. non-cumulative) culture. I then examine the extension of the notion and the variety of referents it may include, which, despite being widely accepted and fairly unproblematic in the "plain" version, may represent a problem for the derived notion of cumulative culture.

The focus on evolutionary perspectives means that I will largely ignore the alarming number (164!) of definitions of culture listed by the oft-cited Kroeber & Kluckhohn (1952) study, which includes other discipline-specific and folk definitions, and concentrate on a suitably limited subset of accounts. The definitional market in evolutionary studies is also fairly turbulent, ²⁶ but convergence is more widespread and more revealing than disagreement.

The evolutionary definition of "culture" is intentionally broad. It is by no means restricted to humans, and it largely lacks many of the everyday connotations of the term, which habitually evokes scholarship (as in "She is a woman of widespread culture") or artistic, social, and ritualistic customs (as in "Maori culture is fascinating"). Table 2 lists a brief selection of standard definitions of "culture", proposed by influential players in the cultural evolution debate.

²⁶ For example, debate has reached a chronic stage concerning whether a broad or narrow definition would be preferable, whether the definition should be modelled on the everyday usage of the term or on a more scientifically construed one, or whether it should be anthropocentric or not (Laland & Galef, 2009; Laland & Hoppitt, 2003).

Author	Definition
Hoppitt & Laland, 2013: 4	Cultures are those group-typical behaviour patterns shared by members of a community that rely on socially learned and transmitted information.
Richerson & Boyd, 2005: 5	Culture is information capable of affecting individuals' behaviour that they acquire from other members of their species through teaching, imitation, and other forms of social transmission .
Jablonka & Lamb, 2005: 205	Culture is the system of socially transmitted patterns of behaviour , preferences, and products of animal activities that characterise a group of social animals.

Table 2: Evolutionary definitions of "culture"

Though wording may differ to some extent, these definitions (along with many others) clearly converge on a minimal set of common elements. The most important are probably *behaviour* and *social transmission*. Further common elements are *typicality* (i.e. behavioural traits shared by most members of a population), behavioural *products* (e.g., artefacts) and cultural *information*. I will briefly expound on these below.

Everyone agrees on culture being dependent on *behaviour*. So all culture requires behaviour; but not all behaviour is cultural (i.e. behaviour is necessary but not sufficient for culture). To be called "cultural," behaviour needs to have spread and been established by means of socially mediated learning (Laland & Galef, 2009). A behavioural trait acquired by means of asocial or individual learning (imagine trial and error learning of predatory tactics to approach

unfamiliar prey, for example in spiders) or controlled by genetic and developmental factors (imagine courtship or distraction displays, such as injury-feigning, in many birds species) cannot be classified as cultural, even if it shows typicality. The very same behaviour—say, predatory tactics or injury feigning could become cultural *if* it were acquired via social interaction rather than through asocial learning or maturation. So there is nothing about the formal properties of behaviour *per se* that makes it cultural or non-cultural: it's the modality of acquisition that matters. An interesting consequence of this fact concerns the "burden of proof" required to detect cultural behaviour: it's impossible to tell cultural from non-cultural behaviour by simple observation of performance—we need to document the *process* of acquisition.

Typicality is another important feature: for many authors, the diffusion of a behavioural pattern in a population appears to draw the line between merely socially transmitted and cultural behaviour. However, in contrast to social transmission, typicality is sometimes argued about (cf. Laland & Hoppitt, 2003): where shall we place the threshold for a behaviour to count as typical (and therefore cultural)? Is sharing by more than one member of a group (say, merely two) enough? Does the trait have to be adopted by the majority of the social group members, or is sharing by a subset enough?

I think that the debate on typicality is misguided. What seems to distinguish *cultural* from merely *socially learned* behaviour is *repeated social transmission* rather than typicality. Typicality is merely a common (but not necessary) consequence of repeated transmission under standard conditions, i.e. in a group where individuals overlap generationally and learn horizontally from each other. To see why we should favour repeated transmission over typicality in the definition of culture, consider a hypothetical case of fully private, vertical transmission: B learns behaviour X from A; C learns it from B; D from C, etc., perhaps under secrecy conditions. Assume also that agents A, B, C, D... don't overlap consistently in time (except for the time needed to learn the behaviour) so that at each time only one individual in the population possesses

behaviour X. X certainly doesn't display typicality; nonetheless, I think most people would be inclined to think of X as a cultural behaviour. In fact, some realworld cases of ritual or esoteric knowledge may work in a similar way: shamans pass on their secret knowledge only to their entrusted successor, who will do the same once he has acquired the status; nobody would seriously deny cultural status to shamanic knowledge, despite its not being shared by most members of the population. By contrast, if we limit the transmission chain to only one node (A passes on X to B—subtracting the condition of repeated transmission), we seem to lose cultural transmission and obtain simply a social learning event. (Morin shows a similar intuition in considering long and narrow transmission chains as the paradigmatic examples of culture, see Morin, 2015.) In sum, I think that typicality is often taken to be a necessary attribute of culture because it is the ordinary effect of repeated transmission under standard conditions in a group, but it should not be considered a necessary element in the definition of "culture".

If the main ingredient of culture is socially transmitted behaviour, most evolutionary definitions of culture include two other types of entities that are strictly correlated to behaviour but are hardly classifiable as such. These are products of behaviour and (cultural) information.

Let us start with products of behaviour. The prototypical behavioural product is an artefact. A clay pot or a defoliated termite-digging stick are good examples. These objects are cultural to the extent they are the result of a socially acquired skill of pot-making or termite-digging. But there may be more complex, abstract examples of cultural products: language (spoken and written) and institutions (e.g., the Eastern Orthodox Church or the DExEU, the British Department for Exiting the European Union) may be listed as example of more insubstantial cultural products.

Culture is generally thought to include another necessary, less corporeal correlate of behaviour. A pot-maker certainly holds a representation of how to make pots: the various phases of modelling, drying, scraping, and firing, along

with the knowledge required to perform these actions in the correct way (what is the optimal heat in the pit or kiln, how long the drying takes, how to scrape thin the dried piece without breaking it, etc.) will be stored in his brain in the form of mental representations (or perhaps in his body in the form of motor programs). Such representations, to the extent they have been socially learned and transmitted, are generally classed as cultural *information*. But cultural information may not only be encoded in minds and bodies. The research trend known as "distributed cognition" suggests also that artefacts and more generally behavioural products represent "storage devices" for cultural information (Cole & Engeström, 1993; Sutton, 2006). Thus a finished pot can carry information about its making; written cultural products such as books typically contain all the information needed to recreate a copy of them and more (they are, in fact, easily copied).

Information deserves a special mention because it is often thought to be the main component of culture, or indeed to *coincide* with culture: on some accounts, culture is no different from (repeatedly) socially transmitted information. Boyd & Richerson's and Hoppitt & Laland's definitions in the earlier table are examples of this way of thinking; Sperber and his associates (aka the cultural epidemiologists) also take "ideas" to be the main material of culture, and encourage us to think of it as a set of public (physical) and private (mental) representations. The idea of the Little Red Riding Hood story that I have in my head is a private representation; the version I tell my nephew at bedtime or the printed edition you borrow from the library are two of its countless public representations (Sperber 1996: 24; more on this in Chapter 4).

The appeal of the notion of information appears to stem from its promising a streamlined treatment of the variety of substrates encompassed by the notion of culture, involving organisms, objects, incorporeal mental representations and social institutions (Jablonka, 2002b; Levy, 2011; Richerson & Boyd, 2005; Sperber, 1996). But this comes at the cost of significant abstraction and unclear benefits. It is beyond the scope of this work to establish whether the notion of

information can provide an effective, seamless bridge among the diverse entities we appear to consider as "culture". In general, however, it should be noted that even authors giving in to the temptation of informational "common currencies" maintain (in one way or another) the distinction between behaviour, its product, and its internal representations. This depends, I think, on the fact that different substrates behave differently, for example when it comes to transmission potential and transmission dynamics. Mental representation, for one, seems in principle non-transmissible unless made apparent in some physical form. (This has often been expressed by saying that cultural transmission cannot be "silent"—i.e. unexpressed but still transmitted—in contrast to, say, genetic information in recessive alleles. This lies behind Sperber's distinction between "private", i.e. available only to the possessor, and "public" representations, available to peers.) Learning how to make an artefact by reverse engineering a model and observing the manufacturing technique are also commonly thought of as rather different processes, potentially resulting in more or less accurate replicas of the artefact (Acerbi & Tennie, 2016; Caldwell & Millen, 2009; Lycett, Schillinger, Kempe, & Mesoudi, 2015; Wasielewski, 2014). In accordance with these observations, I thus maintain the distinction between organismal, physical (inanimate) and mental substrates of cultural phenomena. In sum, notwithstanding the broad consensus on the foundational components of social transmission and behaviour, "culture" seems to legitimately designate a rather miscellaneous class of entities and phenomena.

But there are two further important sources of ambiguity in the notion of culture: I label these "level" ambiguity and the "domain" ambiguity. Let us consider levels first. All the criteria mentioned so far (social transmission, repeated social transmission and typicality) presuppose an extra-individual dimension. To put it in a slogan: it takes two (and plausibly more) to culture. Thus culture has an inherent *collective* dimension. More precisely, the notion of "culture" legitimately applies to single traits, i.e. instances of socially transmitted behaviours, artefacts or pieces of information (for example, a single

115

occurrence of socially transmitted nut-cracking behaviour, or to a particular clay pot) as well as to larger sets of cultural entities, such as repertoires in populations, groups or species (as in "chimpanzee culture", denoting the whole set of socially transmitted behavioural practices in a group). Thus the foundational notion of plain culture is a *level-neutral* notion, equally designating objects and phenomena at the local/individual as well as the global/collective or populational level. The ambiguity about levels is important. I will suggest that it underpins the clearest case of competing and irreconcilable definitions of cumulative culture, by obscuring important distinctions in the foundational notions of accumulation and complexity.

Finally, human culture (and to some extent animal culture) is often analysed in terms of different domains—broadly speaking, areas of functional significance of a cultural trait. For example, it is not uncommon to read of a technological cultural domain as opposed to a ritualistic or symbolical domain in the case of humans; for (nonhuman) animals, "technological" traditions in the foraging domain (such has honey-dipping in chimpanzees or sponge-wearing in dolphins) may be contrasted to "non-technological" social traditions (such as for example the "eye-poke" social convention in white faced capuchin monkeys)although the term "technology" is here more relaxed. (For examples of domain distinctions, see Dean et al., 2014; Fragaszy & Perry, 2008; Perry, 2011; Whiten & Van Schaik, 2007.) For simplicity, I will talk henceforth of a "technological" domain as opposed to "non-technological" one, or for brevity "tech vs. nontech" (which seems the most widespread and general distinction). The tech vs. non-tech distinction is informal and often not explicitly motivated, but generally the former domain is said to include things like tools, tool manufacturing, skills (foraging, shelter building, navigation, etc.), procedures and techniques (fishing, hunting, resource extraction etc.); while the non-tech domain includes social and symbolical traits like body art and art in general, tales, myths, rituals, religious and spiritual practices and beliefs, decoration, style, dance, singing, and so on.

I will also argue that the distinction between tech and non-tech is crucial, for it helps to explain the fundamentally different accumulation dynamics and concepts of function in cultural traits belonging to either domain (section 3.7). Much confusion derives from failing to systematically adopt the domain framework and from speaking of "general" cumulative culture (Morin, 2015; Sterelny, 2017). The distinction between tech cumulative culture and non-tech cumulative culture should instead be embraced, especially in the context of RA explanations. It may even ultimately provide a more principled justification for the widespread but casual talk of "domains".

In sum, plain culture is at the same time conveniently minimal (eliciting consensus) and inconveniently broad, encompassing such a large variety of referents that ambiguity almost certainly looms ahead. While there is no real controversy in thinking about behaviour, objects, collective/individual and tech/non-tech cultural traits as *culture*, when it comes to establishing whether any such thing is or isn't "cumulative", things become rather more challenging. My opinion is that such distinctions matter considerably, acting at times as a potential source of ambiguity and whose delineation will help clarify some apparent paradoxes in the definitions of "cumulative culture".

3.3. Historical notions of cumulative culture

The idea of cumulative culture, intended loosely as showing progressive incremental modifications, is not new. Scholars working in history of technology have long had a sense of progression of technological achievements and of gradually changing morphologies of artefacts that appear to yield ancestry relationships. One early and explicit example of this way of thinking about culture is to be found in the work of Henry Balfour and Augustus Henry Lane-Fox Pitt-Rivers.²⁷ Balfour, first curator of the Pitt Rivers Museum and prominent archaeologist and ethnologist of his times, collected and published a number of essays by Pitt-Rivers (the donor of the rich and diverse ethnographical and archaeological which initiated the Museum at Oxford University). The collection of essays was tellingly titled *The Evolution of Culture and Other Essays* and capped with a preface by Balfour, where he aimed to detail "the first-fruits of the earliest systematic attempt to apply the theory of Evolution to the products of human handiwork" (Balfour, 1906: iii).²⁸ In Balfour and Pitt-Rivers' idea, the organisation of artefacts follows not only local and regional criteria, but a "phylogenetic reconstruction" of artefacts' lineages can be accomplished, based on their morphology, functional destination and distribution, as shown in some of the book's tables:

²⁷ Although an even earlier and equally stark example of "Darwinization" of culture appears just over a decade after Darwin's *Origin of Species*, from the hand of Edward Burnett Taylor—widely considered the founding father of cultural anthropology. His 1871 monograph *Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art, and Custom*", unlike Balfour's work, is more concerned with the symbolic and ritual domain of culture.

²⁸ The Pitt Rivers Museum is interestingly one of the few (or perhaps the only) ethnographical and archaeological museums to organise its collection into "typological displays", where objects are arranged according to types and functions rather than geographical/ethnographical provenience. The idea is due to Balfour himself and to the initial collector, Lane-Fox Pitt-Rivers, whose penchant developed out of an early professional interest in the history of firearms. Typological displays intend to show precisely how different people and societies "solved the same problems" in different times, and how this reflects in the different objects' designs.



Figure 13. Artefacts morphologies yield gradual function-based lineages linked by ancestry, in Pitt-Rivers work collected by Balfour

In his preface, Balfour highlights not only the broad progression of material culture to more efficient forms through series of subsequent modifications, but explicitly applies the evolutionary framework from biology to culture, observing striking parallels (and divergences). In his view, the "innate conservativism of the human species" (ibid., xiii) explains the existence of such "inefficient and yet so elaborate instrument" as the harp of ancient Egypt, Assyria, and India, which can be thought of only as preservations of pre-existent, inherited formal stages.

On the other hand, Balfour notes a particularly interesting dissimilarity between cultural and biological evolution: the lack of linearity displayed by artefacts lineages. This form of culture that we may call *combinatorial* (rather than cumulative), and which Balfour christens "hybridisation of ideas", represents in his view one of the most important engines of technological advancements.

This process of grafting one idea upon another, or, as we may call it, the hybridisation of ideas and experience, is a factor in the advancement of culture whose influence cannot be overestimated. It is, in fact, the main secret of progress. In the animal world hybridisation is liable to produce sterile offspring; in the world of ideas its results are usually far different ... [H]e who would trace out the phylogenetic history of any product of human industry will speedily discover that, if he aims at doing so in detail, he must be prepared for disappointments. The tangle is too involved to be completely unravelled. The sequence, strictly speaking, is not in the form of a simple chain, but rather in that of a highly complex system of chains. (ibid., xii-xiv)

In sum, the idea of historical and cumulative processes at work in culture has been well represented at least since the previous century, and perhaps beyond, and the suggestion of parallels (and relative divergences) with biological evolution were advanced almost immediately after a theory of biological evolution was available. Balfour's account in particular is striking in stressing all the crucial elements that are generally mentioned in modern RA of cumulative culture definitions. These may be summarised as follows:

• Ancestry or inheritance. Similarity between artefacts is explained by reference to inheritance, much as in biological population parent-

offspring relations which predict the population's state at generation N_1 from the population's state at generation N. Ancestry also explains puzzling features like the retention of unnecessary traits (like the inefficient harps mentioned above) and the tree-like appearance of lineages (cf. ibid., xii);

- Gradualism. Artefacts' forms change in time through small modifications,
 i.e. in a gradual rather than abrupt fashion—a feature that Balfour
 ascribes to the "difficulty experienced by man, in a primitive condition of
 culture, of emancipating himself from the ideas which have been handed
 down to him, except by a very gradual and lengthy process", and that he
 also associates with a process akin to blind variation in biology ("causing
 him to exert somewhat blindly his efforts in the direction of progress",
 ibid., xii-xiii).
- Progression or directionality. Successive changes to artefacts are understood as ameliorations or improvements, particularly directed at increasing the "efficiency" of the object (cf. ibid., xii).
- Combinatorial culture. As mentioned before, improvement is obtained not only along single lineages, but also by mixing and matching different types of artefacts. For example, Balfour suggests that stringed instruments derive from adapting hunting bows to musical tasks (see 3.7.1). This feature, which depends on the horizontal flow of information in cultural settings (and that he understands thus as abnormal, intraspecific "hybridisation", a departure from the biological model), has often been neglected in the modern discourse on cumulative culture. However, it has recently made a noteworthy comeback, although under different labels (cf. for example Muthuskrishna and Henrich's idea of the "collective brain", where information travels horizontally in a community,

often across domains, producing innovation through recombination, serendipitous accidents, and cooperative sharing, Muthukrishna & Henrich, 2016).

Thus early notions of cumulative culture are very similar to modern RA-oriented ones; they differ, however, in having a less pronounced theoretical commitment to the evolutionary features of the definition. The hundred years or so separating Balfour's speculations from the advent of the RA have seen a succession of attempts to apply evolutionary ideas to culture, and the bar has risen, making the standard for the evolutionary analogy more stringent. If Tomasello's ratchet effect was really intended to be an informal re-description of cultural change via biological analogies, it has not been received as such. As I emphasised in chapter 1, subsequent readers have certainly understood it as an *argument* for explaining the emergence of cumulative culture in an evolutionary setting, which cried out for testing. The necessity for an empirically tractable definition of "cumulative culture" has grown accordingly.

The first steps in this direction involved dropping the inherent anthropocentric bias of example-based, informally analogical descriptions of cumulative culture. Efforts have moved in this direction, attempting to reformulate the definition as a set of analytically exhaustive and objective criteria (easing the centrality of human-based examples) and taking the analogy with biological evolution more seriously. A definition based on criteria affords generality and can be used comparatively. In addition, the comparative perspective can be seen as stretching into the past rather than across existent species, to evaluate the behavioural production of extinct hominin species and lineages—with the crucial question changing to when, rather than where or whether, cumulative culture appears.

The job of working out the right criteria was left to those who picked up the legacy of the RA in the following twenty years or so, translating it to a

122

decidedly evolutionary framework. I now turn to discuss contemporary criteriabased definitions.

3.4. Contemporary notions of cumulative culture

Modern definitions of "cumulative culture" have generally attempted to move from examples to criteria, in order to operationalise the definition and transfer it into a genuine evolutionary agenda. The result of these efforts is however rather perplexing at first sight. The common goal is to account for the increasing complexity and efficiency in culture through accumulation processes, but what we observe is a conceptual landscape fragmented into a number of apparently disparate operational criteria. Thus, for example, various authors support their claims of cumulative culture in nonhuman animals or ancient hominins by appeal to criteria such as trait *diversification* (the existence of numerous variants of the same object or behaviour, cf. Hunt & Gray, 2003; Whiten, Hinde, Laland, & Stringer, 2011); trait composition (objects being composed of more than one element, such as hafted spears or bows, cf. Lombard, 2016; Muthukrishna & Henrich, 2016; Stout, 2011; Wadley, 2010, 2016); trait pre-planning (the fact that an object or behaviour needs to be "thought ahead" or shows evidence of a mental model to conform to, rather than being performed spontaneously-for example, bifacial Acheuleans; a similar criterion is that of "curated" as opposed to "expedient" tools, famously due to Binford, see Binford, 1980; Kuhn, 1992; Stout, 2002). All these criteria certainly make a connection to our intuitive idea of cumulative culture, but they often represent secondary effects, distinctive but unnecessary features (like typicality for plain culture, for example)corroborative evidence rather than definitional standards for cumulative culture.

For example, observing several "models" of the same kind of artefact in a population has been interpreted as diversification and held as a criterion for

cumulative culture detection. Diversification *generally* signals cumulative culture, as in Balfour's tables, because different forms are interpreted as modifications deriving from an original artefact. But especially with ancient or simple nonhuman technology, we have no guarantee that what looks like "diversification" is in fact just a case of "independent innovation" (i.e. "variants" are not modified versions of an original, but have been invented separately and just *happen* to be similar—often because the scope of innovation is limited and tends to fall back on comparable materials and techniques). Hunt and Gray's claim of cumulative culture in Caledonian crows based on the production of wide, narrow and stepped Pandanus tools (depicted in Figure 1) has been revised in just such sense (Hunt, 2014).

The other criteria have similar problems, leading to applicative difficulties particularly in the detection of minimal and liminal (i.e. early) stages of cumulative culture. The reason for this may be that they lean on the "rich" side of the definitional spectrum of cultural phenomena, incorporating features that are well represented in massively cumulative human culture, but may be missing in simpler instances.

More generally, claims of *minimal* cumulative culture face what may be called a "burden of proof" problem (as Dean and Laland have recognised, Dean et al., 2014). As much as we have no principled way to tell non-cultural from cultural behaviour from its surface properties but need instead to witness its process of social acquisition, we can claim that a trait is cumulative only when we observe its stepwise modification *process*. Especially for liminal cases of cumulative culture, synchronic signatures of the process are unavailable or too ambiguous. Things are different in the case of rich, fully-fledged modern human technology: for these, we can often rely on a variety of sources that exhibit unmistakable historical accumulation; it is often obvious that the trait could not have emerged otherwise.

We might thus want to dispense with secondary criteria like diversification or composition and focus on the central notions of accumulation, complexity

124

and efficiency that seemingly underlie all discourse on cumulative culture. Even on this side, however, the situation is problematic. In the following section, I discuss two competing accounts of cumulative culture that *purport* to define cumulative culture unambiguously as a process of accumulation of modifications that yields greater cultural complexity and improvements. However, as it will immediately be clear, these accounts describe rather different phenomena. The paradox is due primarily to the neutrality of the term "culture" about levels mentioned in section 3.2, but also and crucially due to adopting non-matching notions of accumulation, complexity and efficiency. The two accounts end up producing non-coextensive definitions of cumulative culture (i.e. they pick out different referents in the world), which in turn call for fundamentally different explanatory paths in the RA.

The following section presents these two competing accounts, which I identify initially as cumulative culture as *accumulation of traits* vs. cumulative culture as *accumulation of trait modifications*, and roughly attribute respectively to Jablonka and Tomasello. I then proceed to disambiguate the subtly but importantly distinct notions of complexity, accumulation and efficiency that the two accounts rely on.

3.5. Two competing definitions of cumulative culture

This section focuses on the two clearest cases of competing definitions of cumulative culture.²⁹ I first give a few examples of applications of their central criteria and then show how they lead to irreconcilable pictures of cumulative culture. Both accounts revolve around the idea of accumulation processes in culture; in the absence of better labels and due to the immanent terminological ambiguity, I label these two accounts as follows:

²⁹ For the contrastive identification of these two notions of cumulative culture, see Enquist, Ghirlanda, Jarrick, & Wachtmeister, 2008; Dean et al. 2014; Tennie et al. 2009.

- *Cumulative culture as accumulation of traits* (championed by Jablonka and Avital and endorsed by a few other authors)
- Cumulative culture as accumulation of trait modifications (closer to Tomasello's original interpretation, and by far the most widely used definition).

For simplicity I will mostly be referring to them respectively as Jablonka's and Tomasello's accounts of cumulative culture, from their first or firmest proponents. Although Tomasello's interpretation is certainly mainstream, Jablonka's account is extremely important for understanding the different and often implicit use of the notions of complexity and accumulation in the wider debate on cumulative culture. (In reality, the account is due to Avital & Jablonka, 2000, but given that Jablonka has defended these ideas more consistently and specifically within the cultural evolution debate, I signpost the account with her name.) Other authors may not embrace the "full package" of ideas that Avital and Jablonka present as a unitary definition (or fail to acknowledge it), but as a matter of fact they will often select and adopt the same notions of complexity and accumulation that this account is based on.

3.5.1. Jablonka's account: accumulation of traits

The clearest usage and defence of the first criterion—cumulative culture as accumulation of traits—comes from the work of Avital and Jablonka (2000). The authors treat the criterion as definitional, supporting unambiguous identification of cumulative culture in at least two nonhuman species, macaques and black

rats.³⁰ Although they are not alone in using the criterion to defend claims of nonhuman cumulative culture (see Boesch, 2003 for chimpanzees; and van Schaik, 2003 for orangutans), they make by far the boldest claim.³¹ In particular, Avital and Jablonka rely on a famous series of studies by Japanese ethologists on Koshima Island describing the development of new local habits in a troop of *Macaca fuscata* after the introduction of a new source of food by researchers (Kawai, 1965; Kawamura, 1959). The process of accumulation according to Avital and Jablonka can be described as follows:

- (i) artificial introduction: researchers scatter a batch of sweet potatoes in the troop's foraging range;
- (ii) innovation: Imo (an especially smart female individual) starts collecting potatoes and washing them in a nearby stream (trait 1: potato-washing);
- (iii) diffusion: other individuals follow;
- (iv) expansion: some monkeys discover that biting and soaking potatoes in the sea "seasons" them and makes them more palatable (trait 2: potato-seasoning);
- (v) artificial introduction: researches scatter wheat on the beach;

³⁰ "The long-term study of Japanese macaques on Koshima, which we mentioned in chapter 1, shows that cumulative cultural evolution in animals does occur" (Avital & Jablonka, 2000: 97).

³¹ Other authors give more nuanced accounts of this criterion and seem to count it as an ancillary rather than defining feature. For example, in recent modelling work, Lewis and Laland indicate the *number* of traits (as well as the number of lineages) as a distinctive aspect of cumulative culture (Lewis & Laland, 2012); similarly, Whiten and Schaik, who represent cultural evolution as a pyramid of increasingly complex (and correspondingly rare) forms of cultural phenomena, seem to have in mind the accumulation of traits criterion for the first steps of the pyramid (Whiten & Van Schaik, 2007).

- (vi) innovation: Imo throws wheat in the sea, quickly separating wheat from sand and avoiding laborious collection of single grains (trait 3: wheat-sorting);
- (vii) diffusion: others follow;
- (viii) expansion: a social habit of frolicking and cooling in the sea in summer develops from enhanced proximity to the sea due to the newly developed foraging habits (trait 4: sea frolicking);
- (ix) expansion: young and hungry males start occasionally fishing in shallow seawater (trait 5: sea fishing).

In Avital and Jablonka's description, the population goes through cycles of innovation, socially mediated diffusion, occasional expansion with further innovation prompted by the newly acquired cultural behaviour. The process conveys a clear sense of *accumulation*: traits are introduced and maintained in the population's repertoire. In addition, according to the authors, cultural complexity is unmistakably on the rise: at the end of the process, the troop of macaques has moved on to a *more complex* lifestyle, comprising a larger array of cultural behaviours (and spanning different domains, i.e. social and foraging). Finally, the introduction of new traits offers the opportunity for further exploration and innovation, suggesting (perhaps) a general sense of *improvement* in the repertoire.

3.5.2. Tomasello's account: accumulation of trait's modifications

Let us look now at an example of the application of the second criterion: accumulation of trait modifications. As mentioned before, this criterion is perhaps the most commonly adopted, thus there is a wealth of examples to choose from. One comes from Tomasello's earliest outlines of the notion of cumulative culture (Tomasello et al., 1993):

Human cultural traditions often show an accumulation of modification over generations (i.e. the ratchet effect). Thus, for example, the history of hammerlike tools shows a gradual increase in complexity over time in human prehistory; the evolution of modern hammers in Western culture up to and including steam-driven pile drivers shows a similar increase in complexity over time (Basalla 1988). (Ibid.: 508)

Although not extensively detailed by Tomasello in the work quoted above, the progressive modification of "hammer-like tools" is one of the favourite case studies in the history of material culture and technology: from Basalla's classic work quoted by Tomasello (Basalla, 1988) to contemporary online repositories of engineering curiosities, percussion tools are shown to have evolved from simple hand-held stones, to hafted implements, to a staggering thirty (and counting!) different models of hammering tools. Example include claw hammers (with claws added to the percussion head for removing nails), framing hammers (with straight rather than curved claws, designed for separating material such as skirting board rather than extracting nails), half-hatchet hammer (featuring a percussive surface and a cutting one, shaped after an axe), and chasing hammers (optimal for metalwork and riveting, with a long, rounded handle for maximum impact and a ball-peen tool used to sink rivets flat). Each of these specialised models boasts the basic structure of a traditional hammer, improved for
specialised tasks through one or more modifications to the basic design (handle length, claw addition and/or claw angle, different shape and texture of the impact surface, etc.).

Again, the history of hammer-like tools shows a clear sense of *accumulation*: the percussive element originally represented by a plain stone is retained and improved with the addition of a handle; the handle-and-head model is retained, and further modified with addition of claws or sharp edge; the claws are retained in some models, modified into a different slant, or complemented with a ball-peen; and so on. Undoubtedly, the tool becomes *more complex* over time, being composed by more than one part. Finally, the addition of handles, claws and the like, makes each particular hammer model "a better tool for the job" (or, more precisely, for a specific job). This last sense of "improvement" appears quite different from Jablonka's one, representing more clearly our intuitive sense of increased *efficiency*. An increase in efficiency thus characterises Tomasello's cumulative culture besides an increase in the scope of culture (demonstrated by the larger number of hammer models, which allow performing more and varied tasks). Let us now compare the two resulting definitions of cumulative culture.

Both Tomasello's and Jablonka's accounts of cumulative culture purport to satisfy the crucial rationales of accumulation: increased complexity and (perchance) directional improvement. Yet we immediately perceive that they describe fundamentally different phenomena. The primary reason for the discrepancy between the two accounts lies in the ambiguity of plain culture about levels. As I said in section 3.2, notions of "plain" culture can refer both to collective entities, such as the whole repertoire of a population, as well as to single, localised trait lineages or behaviours. This allows both Tomasello and Jablonka to claim that "culture" accumulates, becomes more complex and improves: but what becomes more complex and improves, in Jablonka's case, is the overall *repertoire* or *lifestyle* of the troop, while in Tomasello's examples it is *individual traits*. In fact, the discrepancy is even more profound than it looks. By shifting the fundamental object of accumulation from singular cultural traits to whole populations' repertoires, the two accounts of cumulative culture end up relying on different notions of accumulation, complexity and efficiency. Divergence in these foundational notions has gone largely unnoticed and has crept into the discussion of cumulative culture.

Let us briefly look at accumulation first, which is the simplest case. An incongruity in the usage of the notion has been briefly noted in various places (but to my knowledge, never addressed in detail; see Morin 2015: 223-226 for a mention). While Jablonka's account appears to adopt a *quantitative* sense of accumulation, Tomasello's implies a *qualitative* sense of the notion. In the quantitative sense, accumulation simply refers to "staying there", much as dust accumulates if not swiped clear. In this sense, accumulation is tantamount to preservation. If there is any sense of directionality, it is only quantitative: things simply add up. In the qualitative sense, by contrast, accumulation carries a much clearer sense of progress or directionality: things not only increase in number, but become individually *more efficient*. In Jablonka's account, there is no obvious sense in which potato-washing becomes more efficient—i.e. "better", faster, more economical, more methodical—and the idea that the whole repertoire has acquired such qualities seems implausible, unless reinterpreted as an increase in cultural scope—which seems, however, a different quality.

If this is correct, one problem with Jablonka's sense of accumulation is that it is unclear in what sense it makes cumulative culture different from mere traditions, cultural stability and plain culture, which also minimally require traits to stick around. In addition, the criterion of increased efficiency appears farfetched in Jablonka's account; there may be a sense of "more things being done" by the population, but speaking of such a feature as "efficiency" doesn't seem quite right.

In sum, the quantitative/qualitative distinction suggests that the process of accumulation at work in Tomasello's cumulative culture is fundamentally different from simple preservation. Just *how* different, however, doesn't emerge

clearly from this way of disambiguating accumulation. I suggest that to understand the root of the difference, and account for directionality and progress in cumulative culture, we must turn to examine *complexity* and *efficiency* in cumulative culture: I do this in the following sections.

3.6. Complexity

Complexity is clearly central to cumulative culture: in fact, it is often used as a straightforward synonym for cumulativeness, with the expressions "cumulative culture" and "complex culture" being used interchangeably. Equally widespread is the opinion that cumulative culture increases in complexity, with this feature often being thought as definitional for cumulative culture (Enquist, Ghirlanda, Jarrick, & Wachtmeister, 2008). In effect, complexity bears a straightforward and intuitive connection to accumulation processes, being a result of accumulation; but as we just saw, accumulation processes can be interpreted as occurring in slightly different ways. The notions of complexity that I examine in this section may be seen as products of differently interpreted accumulation processes, mirroring the ambiguities seen for such notions in Tomasello's and Jablonka's accounts.

Judging complexity seems straightforward enough in ordinary contexts, but beyond everyday usage the notion becomes elusive (Ridley, 2000). This represents a problem, for, as it has been recently noted, "complexity" is widely relied upon in cultural evolution studies, but its meaning may not be so homogenous after all (Querbes et al., 2014). I will sketch the most common existing interpretations in the cultural evolution literature, and connect them back to processes of cultural accumulation at different levels.

3.6.1. Three kinds of complexity

In some cultural evolution work, complexity appears interpreted simply as the *number* and *kind* of parts from which an object, process or phenomenon is composed. In a recent and rather animated debate, the concept has been ascribed to Oswalt (Joseph Henrich et al., 2016; Querbes et al., 2014; Vaesen, Collard, Cosgrove, & Roebroeks, 2016), from his early proposal to measure the complexity of foraging toolkits by counting the number of different types of tools' parts (Oswalt, 1972; 1976). Let us label this first interpretation *Oswalt's complexity*. A recent example of its application may be found in Acerbi, Kendal, & Tehrani (2017) who analyse complexity in folktales, defining it on the basis of three parameters: "the number of tale types, the number of narrative motifs, and, finally, the number of traits in variants of the same type" (Acerbi et al., 2017: 474).³²

On the other hand, an account of complexity *as interdependency* can be contrasted to the one just described. In the same heated debate (and particularly in Querbes et al., 2014) complexity as interdependency has been credited to seminal work by Simon (Simon, 1962).³³ This idea of complexity,

³² Such an account has rough parallels in biology. In discussing the evolution of complexity, Ridley (2001) notes that organisms' complexity has often been measured in terms of *number* of *different* parts: these can be number of organs, tissues, cells types in a body, or of behavioural patterns, or of stages in life cycles. Ridley recognizes that, as much as in the cultural context, "complexity is an ill-defined term, [that has] no biologically agreed definition"; nonetheless he suspects that "most people would agree on what should contribute to it", Ridley 2001: 28. He also considers accounts of biological complexity based on the notion of information ("number of bits of information in a body" or "how long it would take someone to describe the life form completely", p. 29), and finally the one he ultimately favours, i.e. the number of genes.

³³ It has been claimed that Simon's account of complexity is "largely concerned with the central notion of hierarchy" (Henrich et al., 2016). While this is exact, it should *not* be intended as a claim that complex systems *need* to be hierarchical in order to be defined as complex, but rather as the claim that complex systems tend to *evolve* towards hierarchical architectures for adaptive reasons. In

which I will call *Simon's complexity*, has been developed in various places and with various tools.³⁴ One particular effective way of describing this type of complexity is through so-called NK models, first developed by Kauffman and Levin to explore the evolution and fitness alterations of genomes (Kauffman & Levin, 1987). In these models, which have been explicitly applied to the evolution of technology, N represents the number of elements in the system and K their interconnectedness, interdependency or epistasis. According to this account, thus, complexity *presupposes* numerosity and heterogeneity, but the emphasis is crucially placed on interconnectedness of parts. As we will see in a moment, interdependencies afford a qualitative dimension to such complexity, which contrasts with the purely quantitative character of Oswalt's interpretation.

Finally, there may be a third sense in which complexity is used in association to cumulative culture. We typically say that some tasks are more complex than others: building a Lego spaceship is more complex than building a plain Lego wall. In this sense, complexity has an epistemic flavour. Complex cultural traits are in this sense *cognitively challenging* traits: they impose a toll on memory, understanding and/or coordination. Even at first blush, this *epistemic complexity* appears associated with the primitive, ontological complexity captured by the two previous accounts: complex cultural traits (made by many different parts, and/or interdependent parts) are likely to be harder to learn, remember, understand and perform correctly. I will pick up on epistemic complexity later on; for now, let me focus on Oswalt's and Simon's notions, and on their relation to the accounts of cumulative culture mentioned so far.

other words, complex systems are more evolvable if hierarchically organized, therefore more feasible and successful—and often more *understandable*, suggesting that hierarchical organization may be in some cases an epistemological rather than ontological feature of the world.

³⁴ This notion of complexity is loosely connected to the literature on complex biological systems, such as developmental systems, and to the notion of holism in biology and social sciences, which I avoid touching upon for their alarming vagueness.

A crude, initial take on the matter is that Jablonka's account of cumulative culture utilises Oswalt's complexity as sheer number of different parts, while Tomasello's is primarily based on Simon's complexity as interdependency. Clearly, Oswalt's complexity has a straightforward connection to the sense of quantitative accumulation associated with Jablonka's account of cumulative culture. If accumulation is simply preservation, then we will observe an increase in number of cultural traits and conceivably in their heterogeneity over time—e.g., as exemplified by the obsessive-compulsive disorder known as hoarding. Notice that Oswalt's complexity leaves relations between accumulating parts completely out of the picture, being therefore compatible with additive juxtapositions implied by quantitative accumulation.

But if quantitative accumulation justifies the claim that the cultural repertoire of Koshima Island's macaques becomes *more complex* over time (they become Oswalt-complex, by simply exhibiting more traits of different kinds than the original one), it doesn't seem enough to account for the complexity of individual cultural traits. New parts, modifications and innovations in an individual trait do not add up in a shamble: you cannot just throw together a handle and a hammerhead and obtain a functional hammer. In this respect, Simon's richer notion appears to do better. The fundamental discrepancy between the two notions of complexity lies in the role of interdependencies in the picture, which introduce efficiency and a sense of qualitative directionality in the evolution of complex cultural traits. To see in what way exactly the qualitative dimension enters the scene, however, we will need to proceed by steps.

3.6.2. Understanding interdependencies

Involving numerosity and heterogeneity, Simon's complexity can be seen as a product of (quantitative) accumulation too; in addition, however, it requires that

accumulating elements do not pile up randomly, but establish meaningful connections: interdependencies. Admittedly, there is often little explicit indication as to how interdependencies should be understood, but they seem to have two main properties. First, interdependent elements are described as "affecting each other" when changed (Frenken, 2006; Kauffman & Macready, 1995; McCormack, Hoerl, & Butterfill, 2011). In accordance with this statement, I will propose that, especially in technological systems, interdependencies may be interpreted as a form of physical, causal connection among parts. (This feature of Simon-complex cultural traits may have important consequences on their cognitive manageability; but I will talk about this later.) Second, interdependencies have another fundamental property: they are *teleologically* organised connections. By describing interdependencies as "teleologically organised connections", I mean that they are systematically arranged and structured to obtain a goal, or better yet to perform a *function*. The best way to understand these features of interdependencies in a complex system is to observe an example of their application from up close: I will do so by describing a case of NK models applied to technology.

As I said before, NK models have been transferred from biology to other fields concerned with the study of complex systems' evolution and optimisation—including economics, management and (crucially for my discussion) technology (Frenken, 2006; Henrich et al., 2016; Kauffman & Macready, 1995; Solée et al., 2013). In these models, N defines the number of elements in the system (with each N_i element being in one of many possible states), while K is the number of other elements in the system to which each element is connected (through "epistatic relations"): in other words, K measures the degree of interaction, interconnectedness or "epistasis" in the system, being a direct function of its complexity.³⁵ NK models were originally conceived to

³⁵ Kauffman's formulation takes specifically into account only those architectures where each element is affected by the same number of other elements, so that NK models describe systems with N elements in which each element is affected by K other elements—but other architectures are

explore fitness alterations in complex biological systems; when applied to technology, fitness is conceived as "efficiency", "quality", or more generally "performance" (Frenken, 2006). (Incidentally, this alone indicates the presence of a qualitative dimensions.) In the words of Kauffman, NK models are "tunably rugged fitness landscapes" (Kauffman & Levin, 1987) representing the system as an abstract multidimensional space (more precisely: a hypercube, as in Wright 1932), where changes in fitness or performance can be visualised as differential heights ("peaks" as fitness optimums and "valleys" as low-fitness zones) and manipulated by intervening on the N and K parameters. The central idea is that the more parts are interconnected (high K), the harder it is to change anything in the system without compromising fitness (i.e. performance). Roughly speaking, highly interconnected systems are more fragile, because even controlled modification of a part may bring about a cascade of uncontrolled effects on other parts.

The NK literature generally focuses on fitness changes; it explores statistical properties of search spaces, and whether and when innovation is better achieved through myopic search (trial and error search in the vicinities of performance peaks) or insightful search (attempting to "hop" from peak to peak, which may yield greater benefits but also bring exposure to risks of performance disruption). The overall goal is to maintain or increase the system's "fitness", or good performance.

Thus Frenken and Kauffman suggest that a car, for example, can be thought of as a system composed by a number N of different, highly interdependent parts or elements: engine, brakes, transmission chain, steering device, etc. (The same can be said of an airplane, or any complex artefact or

possible. For example, each element in the system may have different levels of interconnectedness (non-K), and connections may be measured along a continuum (weak vs. strong connections) rather than counted discretely as present vs. absent. NK systems may thus range in a continuum from fully decomposable (no epistasis = minimum complexity) to fully epistatic (each element affects all other elements = maximum complexity).

procedure like industrial production lines—but cars seem to be a favourite example.) Each element may have two possible alternative design options (say: two ways of implementing brake technology; two mechanisms for steering; two different engines etc.). The total number of possible designs for such car would be 2^N, representing its "design space" (the hypercube). Car designers will move through the hypercube space, manipulating elements (or better: strings of interconnected elements) and searching for better performing options.

I will discuss fitness and performance later; what I want to focus on first is the causal nature of interconnectedness in artefacts. Consider wanting to modify one car element: the transmission system—you want to go from automatic to manual. In this case, presumably, you will have to change the clutch linkage box and pedals; but you may omit modifications to seatbelts or window screens. (I am on treacherous ground here, because I don't really know how cars work; but you get the idea.) The key question about the causal nature of interdependencies, then, is *why*, under this circumstance, you can omit changing screens and seatbelts but *must* change pedals and clutch linkage box—or else pay the price of obtaining a poorly or non-functioning car.

The answer (or a reasonable approximation of the answer) is that these elements are *physically* and *causally* connected. The reason you have to change clutch linkage box and pedals is that such elements are tightly causally linked to the transmission device; while seatbelts or window screens have a negligible to null connection to it. A similar lesson may be drawn from hammer-like tools. Changing from a handle-less to a handled design will induce changes in the percussion head shape: most notably, in common hammers, the percussion head goes from a single block to a perforated solid that allows insertion of the handle bar. Thus, according to the standard characterisation, handle and percussion head are highly interdependent (because changes in the former will necessarily result in changes in the latter). But this is ultimately justified by the physical and causal connection between the head and the handle. Their physical connection ultimately causes the handle to work as a force amplifier for the impact surface (by converting mechanical work into kinetic energy).

Talk of physical connection and causal influence, however, is not enough to explain the difference between Oswalt's and Simon's complexity. Intuitively, we judge a car as more complex than a pile of car parts (composed of the same number and types of elements), and Simon's complexity explains this intuition by appealing to interdependencies between car parts. If interdependencies were simply causal/physical connections, however, consider placing all the car pieces in a giant tumble dryer: certainly now the parts are causally interacting in a very complicated way. However, we still don't seem to get the same sense of complexity that we associate with the assembled, *functioning* car. Why? Interdependencies do not only describe causal connections, but *normatively regulated* causal connections—where normativity in turn derives from the *teleological organisation* of the complex (*sensu* Simon) system.

In a teleologically organised system, parts are not thrown together at random (i.e. they are not simply numerous and heterogeneous) but are assembled in a specific way *in order to* achieve a goal, or a *function*. A car's parts need to be arranged in such and such a way *in order* for the car to work. Shuffling together a handle and a hammerhead (such that they are in causal interaction with each other) doesn't give you a functional hammer, but the head needs to be attached at the tip of the handle *in order* to amplify the impact force. A genome is composed by an intricate, ordered, and teleologically structured network of epistatic relations *in order* to produce a viable phenotype. The *normative* dimension of causal connections imposed by teleology means that these connections may be "right" or "wrong", i.e. might perform better or worse, or not perform at all, against the normative standard set by the system's function.

At this point, it will be natural to ask how complex systems acquire such function. One may answer such question in different ways: on one hand, function may be naturally seen as imposed by agents (as is normally the case in

technology: a car and a hammer are designed and built in such and such way because we want to drive around and drive nails into hard surfaces). On the other hand, function is often regarded as a product of natural selection processes (or indeed, *the* product of natural selection, inasmuch as adaptations can be conceived as *functional* traits, and functions contribute to organisms' fitness: the function of flowers in plants is to attract pollinators; flowering is a plant's adaptation that greatly contributes to its fitness). This language, though contested, is pervasive and heuristically fruitful in biology, where, however, it must be used without reference to intentional, agential purposes or goals (Ariew, Cummins, & Perlman, 2002; Amundson & Lauder, 1994; Beckner, 1969; Mayr, 1998). But I will examine the different types of function of cultural objects and their relations in the next section.

As one might guess, teleological organisation or function is closely related to the criterion of efficiency that both Tomasello and Jablonka aim to incorporate in their accounts (with different success): efficiency refers to how well a trait performs its function (more on this below). Ultimately, the reason Jablonka's notion of cumulative culture struggles or downright fails to account for the increase in efficiency is that cultural repertoires lack a (tight) teleological organisation and a function (or at least they do in nonhumans). This also results in Jablonka's account's failure to capture qualitative changes in culture. Again, notice that incorporating plain causal connections into the account wouldn't do. The addition of new traits to a cultural repertoire may also be represented as a series of causal connections between traits. For example, in Jablonka's example potato washing leads to or *causes* (in a very general sense) the innovation of potato seasoning. But such causal connections are profoundly different from interdependencies: there is no obvious sense in which they are "right" or "wrong" (normatively regulated by a teleological dimension), and no obvious sense in which they make the repertoire work better or worse. Potato washing might have led to a completely different trait (playing at throwing potatoes, for example) and this would not have mattered; but the casual connection between

potato washing and potato seasoning isn't really "more right" than the one between potato washing and potato playing. This is in clear contrast with the connection between handle and hammerhead, or between car parts, which must be arranged in the right way to afford a well-functioning vehicle or tool.

In comparison, Tomasello's account of cumulative culture, by referring first and foremost to single (and mostly technological) traits, fully exploits Simon's sense of complexity and opens the way to the qualitative dimension of cultural accumulation. Accumulation is said to be qualitative because it conveys a sense of increasing efficiency; the qualitative dimension itself derives from the teleological nature of function—inasmuch as function can be performed more or less successfully, i.e. efficiently.

In sum, it seems reasonable to suggest that the notions of causality and teleology play a role in explaining Simon's interdependency-based complexity, and consequently in notions of cumulative culture that adopt this definition of complexity. Things may not be so simple, but introducing these concepts gains us a foothold in explaining the fundamental difference between quantitative and qualitative accumulation in culture, and between the corresponding competing accounts of cumulative culture illustrated so far. If this is correct, the RA may then have two rather different explananda—and each may require quite a different set of inferences drawn from the basic RA elements. If we take cumulative culture to mean something like Jablonka' definition, high fidelity and imitation will have to explain only the preservation of cultural modifications; but if we take Tomasello's cumulative culture as the explanatory target, imitation and fidelity must also play a role in explaining something like the viability of complex cultural traits, depending on Simon's teleological, interdependency-based complexity. I now turn to examine function and efficiency in more detail.

3.7. Efficiency

While, as we have seen, Jablonka's account of cumulative culture struggles to accommodate the idea of efficiency, the idea seems a central element of Tomasello's account. Many advocates of this account suggest that increase in efficiency is a definitional element of cumulative culture—that is, they suggest that in order to count as legitimate cases of cumulative culture, traits *must* increase linearly in efficiency; the assumption appears integral also to recent attempts to test the RA "in the laboratory" (Caldwell & Millen, 2008a; Dean et al., 2014; Enquist et al., 2008). This claim might have grounds, but needs qualification. As it is, the idea that all cumulative cultural traits increase linearly in efficiency appears to encounter a few difficulties. Let us start by defining "efficiency" more precisely, or at least, by identifying what this term generally means in the RA literature.

A good place to get a first-hand sense of what is generally meant by "efficiency" and its alleged adoption by advocates of Tomasello's account is the history of technology. In addition to the already-mentioned evolution of hammering tools, and to the popular examples of bikes, aircrafts, bridges, etc., consider the simple paradigmatic case of the fork's evolution (or better: of cutlery sets) described by Basalla (1988).



Figure 14. The evolution of forks

Modern forks presumably derive from a single-pronged instrument that was originally used interchangeably to stab food, hold it in place while cutting, and bring it to the mouth. The double and then multi-pronged designs represent modifications and *improvements* that help to keep the food parcel steadier while it is cut, as well as complementing the function of changing types of knifes (hinting perhaps at co-evolutionary processes). The evolution of forks is often said to show an *increase in efficiency* over time, to the extent that multi-pronged versions allow performing the task of "eating neatly" *better* than their singlepronged ancestors. Thus, the notion of efficiency appears straightforwardly to refer to something akin to "better performance" of a cultural trait with respect to a task. We can express this idea more clearly and compactly in terms of *function*. Forks have the function of "eating neatly", and their efficiency measures how well they perform this function.

A few basic qualifications are in order about function and efficiency. Function is clearly a *teleological* but also a *relative* notion; and so is efficiency. Efficiency should always be considered *relatively* to one specified task or function (or more than one, but always specified). It doesn't make sense to speak of "absolute efficiency": forks are quite efficient for eating, much less for brushing teeth; boats are fine to move on water, but rather useless on the ground. How do we determine which function establishes an item's efficiency? I argue that, in general, the literature on cumulative culture (which focuses on tech traits) looks at the task the item is *designed for*, or its designer's function: we may call this "D-function" for brevity.

It is necessary to label D-functions because traits can have other types of function, which may or may not coincide with the D-function and can cause conceptual confusion. One very obvious case is the *user*'s function (i.e. what an object is *used for*): "U-function" for brevity. Someone may choose to use an old hammer as a paperweight. Being old and fragile, the hammer may perform its D-function (knocking in nails) quite poorly, while performing its U-function (holding down sheets of paper) exceptionally well. The D-function of microwave

ovens is to heat food, but a user may decide to dry a little furry pet in one. While the U-efficiency of the microwave will be extremely low in this case, it doesn't generally entitle us to say that microwaves are inefficient.

U-functions are also useful to determine the status of "expedient technologies", i.e. objects that are not literally designed and built by agents, but rather selected and used in virtue of their functional properties-like, for example, a digging stick (Binford, 1980). A stick is a natural object that may be picked up as it is (not built or designed) to dig a hole. Its properties are not determined by an agent, but rather selected from an available set and put to use. If the agents select well, the U-efficiency will be high and will essentially coincide with an ideal D-function. A digging stick has many significant physical properties in common with a designed spade (hard, elongated, pointy, robust), which is why one prefers to get one's hands on a stick rather than on say, a leaf, if the intention is to dig a hole in the ground. So when the agent is well informed, we may presume the U-function of expedient technology will overlap in principle with D-function. When cumulative culture theorists debate cumulative culture's increase in efficiency, they mostly seem to have in mind something like D-efficiency. When the designer/producer and user of a trait are not aligned (i.e. the agent performing or using the trait is not the one who designed it; or when "proper" design is missing like in U-functions), D-function generally sets the standard for efficiency.³⁶ Both D- and U-functions are, in any case, agent-imposed functions.

Aside from task-relativity, there are two other aspects that make efficiency a fundamentally relative notion: its relation to the environment and the cost/benefit trade-off. Relativity to environment is quickly illustrated: clearly, forks work well in a world of medium to soft food, but they would be terribly inefficient in a world of rock-hard bites. As we will see later, technological traits can count on a fairly consistent and predictable environment, so their efficiency

³⁶ Users will be more commonly removed from designers in complex societies with division of labour; less so in simpler societies and in animal cultures.

will generally not be vastly disrupted over time—though importantly, the same cannot be said of the non-technological domain.

"Cost/benefit relativity" refers to the fact that, in general, the efficiency of a trait is evaluated in relation to the cost of the trait to the user or designer. Achievement of performance (and of performance quality) being equal, objects or procedures that require less material, time, investment and energy will be judged *more efficient* than their rivals (i.e. objects and procedures that achieve the same performance with more resource expenditure). In technological traits, "better" will always be evaluated on a case by case basis, taking into account the nature of the task and of the target environment, and the balance of prospective risks and gains. In a slogan, efficiency for technological objects and behaviours might be understood as "better with less".

These primary qualifications aside, the idea that trait efficiency means "better" and "better with less" (relative to task, costs and context), and that it increases linearly over time in cumulative culture, looks fairly plausible at first sight. However, if we take this definition of "efficiency", and maintain that linear increment is a defining criterion of cumulative culture (as, I noted, several authors do) we seem to run into a few difficulties: in certain trait lineages, efficiency seems to decrease or even disappear with the accumulation of modifications. As will be evident, these cases always involve the transition of tech traits into non-tech ones, so they may be understood as the problem of "missing efficiency" in the non-tech cultural domain.

3.7.1. Missing efficiency?

Consider the case of turtle-shell fishhooks traditionally manufactured by Torres Strait Islanders (in Dean et al., 2014). These objects display a curious evolutionary trajectory, being first employed as a subsistence tool, and then modified into ornaments for married women. The hooks currently on display at the Australian museum in Sydney are voluminous, with a broad surface (to allow and enhance the fine decorative carvings); should they be employed for fishing purposes, they would probably do a poor job of it. The fishhook example suggests that the cultural trait has in fact undergone modifications that led to a *decrease* in efficiency over time; even more problematically, it is unclear whether the version modified into an ornament can be said to have *any* efficiency at all. The fishhook case is not an outlier: similar cases are widespread. Balfour offers a nice example, describing the derivation of stringed instruments from hunting bows:

The native of Damaraland, who possesses no stringed instrument proper, is in the habit of temporarily converting his ordinary shootingbow into a musical instrument. For this purpose he ties a small thong loopwise round the bow and bow-string, so as to divide the latter into two vibrating parts of unequal length. (Balfour 1906: vii-ix)

This temporary expedient is perfected by "some African tribes", who manufacture lighter bows for musical purposes only, often resting them "against some hollow, resonant bodies, such as an inverted pot or hollow gourd. In many parts again, we find that the instrument has been further improved by *attaching* a gourd to the bow, and thus providing it with a permanent resonating body" (ibid.: ix). Clearly, while the musical instrument might have improved through such addition, the original hunting bow is rendered useless by these modifications.

The fishhooks and stringed instruments represent cases in which the efficiency of a particular object decreases with its accumulating modification, but more fundamentally they are taken to pose what we may call a problem of *missing* efficiency. It is not at all clear how ornaments, songs, tales, or religious belief can be *efficient* in the sense we considered for forks. If efficiency is really *missing* from non-tech traits, and tech traits habitually transition into non-tech

ones (as seems the case), the claim that cumulative culture increases in efficiency becomes untenable.

Broadly speaking, there are two ways around this problem: (i) abandon the requirement for increasing efficiency in cumulative culture entirely, and expect cumulative culture to show no trend towards greater efficiency over time; (ii) relax the requirement for a linear increase in efficiency in cumulative culture which can be done in several ways. I see (i) as defeatist and an ultimately imprecise solution, which deprives us of important intuitions about cumulative culture. I will thus try to defend a version of (ii). This will require two steps: first, I propose that efficiency is not *really* missing in non-tech traits; rather, non-tech "efficiency" works in such different ways from tech trait that treating them as equivalents is counterproductive and generates confusion. In turn, this will suggest that it is necessary to adopt a double-standard and separate the discourse on tech cumulative culture from that on non-tech cumulative culture. In the next section, I tentatively advance a hypothesis for why tech and non-tech efficiencies behave in such incommensurable ways: this will involve looking at another, more fundamental notion of function and efficiency, and at the kind of environments it performs against. Secondly, I will show that once limited to tech cumulative culture, the claim about increasing efficiency is fundamentally correct, although requiring some qualifications (for example, renouncing the assumption of a linear dynamic of increase).

To see whether non-tech traits have any efficiency at all, we should first ask whether they have functions. How about D-functions? Non-tech traits often seem to lack the kind of purposive D-efficiency that we regularly see in tech traits. In many cases, non-tech cultural traits are not "invented" (designed) by the bearer, but rather emerge gradually and nebulously from existing and developing repertoires. Perhaps non-tech traits have something akin to Ufunctions? This seems somehow more plausible, as in many cases we "pick up" songs, tales, symbols and habits, much as we may pick up sticks for digging. Even so, agents are often unable to clearly articulate *for what* (U-function) they pick up a certain religious belief, model of skirt, or hairstyle, or certainly much less clearly than explaining why they pick up a digging stick or a hammer. If probed, however, many religious people may eventually concede that they adopt a particular religious belief or practice for a purpose—for example, in order to go to paradise, or to look good in the eyes of neighbours, or more generally to feel better about themselves. The same can be said about people who choose a particular skirt type or hairstyle: given thought, they might say that they choose it to look fashionable or to take after their favourite celebrity. So for the sake of the argument, let us assume that non-tech traits can have U-efficiencies and sometimes perhaps D-efficiencies (like when one very *intentionally* creates a particular dress-code or faith).

Even so, these efficiencies seem to "work" in a different way from tech Dor U-efficiencies. In particular, I think we can identify the following discrepancies:

1. Epistemic access.

Non-tech D-functions and U-functions are in general much less epistemically transparent to agents. We are generally pretty good at articulating why (*for what*) we want or build a fork, a digging stick or a car; as just noticed, we seem much more clueless, at least at first blush, about fashion trends, beliefs, and spiritual practices.

2. Lack of reliable feedback.

Even when we can somehow work out a D-function or U-function for a non-tech trait, it is often very unclear whether the trait achieves its purpose (and consequently whether it can be deemed more or less efficient). Even when efficiency seems quantifiable, it doesn't predict trait selection on the part of agents. Clearly, the fact that there is no evidence that being religious does get one into paradise has regularly failed to prevent hordes of new followers from joining in enthusiastically. Compare these cases to the tech examples in the following figure:



Figure 15. Examples of "uncomfortable" (and inefficient) design: the thick fork, the masochistic coffeepot, and the slipper spoon

Here, we are readily able to tell that the thick fork, the slipper spoon, and the masochistic coffeepot (from Norman, 2013) are highly inefficient objects, for they glaringly fail their respective functions of spearing food, spooning liquids, and pouring an inoffensive cup of coffee.³⁷ Accordingly, agents are very likely to quickly discard such objects. If they are kept, it is for *another* function (for example, illustrating the paradoxes of uncomfortable design, which they may do rather brilliantly).

3. No inherent relation between efficiency and structure.

As briefly mentioned in the previous section, tech traits' efficiency has a rather straightforward connection to internal causal interdependencies. Look again at the masochistic coffeepot. Here, the "wrong" physical connection between interdependent elements of the pot (the main body, the spout and the handle) disrupts efficiency of the trait to the point of making it unserviceable. In nontech traits, by contrast, their achieving their D-functions (if they have them) or

³⁷ The "deliberately inconvenient" pieces of cutlery are designed by Athens-based architect Katerina Kamprani. A larger collection can be seen here: <u>https://www.theuncomfortable.com/</u>.

U-functions seems largely disconnected from, or very mysteriously connected to, the structure of the trait. Consider the swastika. You may stitch a swastika patch to your clothes to parade very questionable political views. But, will your political signalling become less or more efficient choosing a clockwise or anticlockwise swastika: $f ext{ or } H$? The answer is no, and explains also why in other places and times swastikas were symbols for the sun, good luck, eternity, and even simple greetings.

Symbols are largely arbitrary, and can be reused, inverted, and associated with anything at any time, losing even the feeblest connection to the original referent (when such connection existed in the first place) through this process of indiscriminate layering and re-use. I think this aspect is particularly important to explain why, ultimately, changing some parts of non-tech traits by imprecise transmission is less likely to disrupt their viability and efficiency. If you copy a dance movement in the wrong way, even if that dance is an important ritualistic element that requires extremely codified movements, the missing or altered movement will not render the whole performance a disaster (or not necessarily). In fact, should you be a renowned dancer, your mistake may even be taken for a respectable innovation (and transmitted as such). By contrast, failing to pedantically copy *every* step in the instructions leaflet of an IKEA piece of furniture (the assemble-it-yourself type) is likely to lead to disaster.

4. No "better with less".

Non-tech traits patently defy the general principle of "better with less" that we found applicable to tech traits. On the contrary, they sometimes derive *more* efficiency from a certain "lavish and deliberate 'squandering' of time and resources" (Dennett 1997). Examples here may involve ritual offerings, architecture, clothing and much more. Consider the Baroque style. This excessively ornate decorative style aims at impressing the viewer, not only at aesthetically enhancing the surroundings. The more Solomonic columns, multi-coloured marbles, gilded leaves and spiralling vines, the stronger the awe-effect

and hence the efficiency of Baroque decoration. But again, it depends. Sometimes minimalism is the goal and rigorous simplicity establishes the efficiency of a non-tech trait (think of Apple products' aesthetics: the simpler, the cooler). There seems to be no clear relation between trait structure and trait efficiency.

In conclusion, even when we can say that a non-tech trait has a Uefficiency or D-efficiency, this seems to be very different from what we intuitively identify as U- or D-efficiency in non-tech traits (and in some cases, it just seems hard to locate at all). I think that such stark discrepancy might be explained by the existence of another more fundamental notion of function to which D-functions and U-functions are differently related. I turn now to this third conception of *function*.

3.7.2. A different kind of efficiency: efficiency and S-functions

I have distinguished D-functions from U-functions, but in both cases the functions are given by agents. While this seems well suited when it comes to artefacts, we commonly speak of non-designed things having functions too: the function of the heart, say, is to pump blood. But no one "gives" heart this function. So, how do hearts have it? A natural (though contested) answer is that hearts have this function because it's their doing so (pumping blood) that explains their existence: they acquire it through *natural selection* processes. By comparison, hearts also make thumping sounds, but that's not their function because it's not by virtue of making this sound that hearts exist. Let's call this kind of function "S-functions." The S-function of the heart is to pump blood, not to make a thumping sound. (On the other hand, since we can use the thumping sound for certain medical purposes, in some circumstances making a thumping sound might be considered a U-function of the heart.) Attached to this notion of

function comes a corresponding type of efficiency: *S-efficiency* measures how well an object performs its S-function (when it has one).

There are parallels and contrasts between the previously mentioned agentimposed D- and U-functions, and selection-established S-functions. Most obviously, S-functions are also, necessarily, relative and context-dependent: mammals' hearts may work well on dry land, but are sub-optimal underwater (Williams et al., 2015). Both types of functions are "teleological", although natural teleology must steer clear of the notion of the intentional goal. This depends on their different aetiology: unlike the intentional design to an object, selection processes are clearly unintentional and must proceed through blind approximations. Although there will be degrees of variation (selective processes may be guided to various extents; many design processes will be trial-and-error like), the idea is that in the first type of process variation is directed and function intentionally imposed, while the second type of process selection is blind and function emerges as a result of selection. Thus, in general, from the perspective of the agent, the epistemic access to function is much more transparent in design-like objects than in products of natural selection, largely because in the first case, function depends on the agent herself or on cognitively similar subjects.

At the cost of crude simplification, there is an important thing to notice about S-efficiency: by definition, it tends to increase (locally, and given constraints) over time. At any time, selection will "keep" comparatively *better performing* (more S-efficient) traits: other things being equal, N₁-generation hearts would not have been "kept" (i.e. have survived) unless they were *better* than or at least equal to N-generation hearts at pumping blood. (Of course, the claim is understood as *relative* to all applicable conditions: environment, physical and developmental constraints, "pleiotropic" traits, drift, etc.). The evolution of the eye teaches the same lesson. Eyes evolved (apparently several times) from primitive eyespots (cell organelles filled with photosensitive proteins) and pigment spots (clusters of photoreceptive cells), to intermediate

"pinhole-eyes", to sophisticated "camera-eyes", with each version affording improvements (and carrying trade-offs) in terms of resolution, colour vision, motion or shapes detection. Thus eyes can be said to have become more efficient relative to particular tasks and environments (and to have been "kept" for this reason). Of course, no biologist in his or her right mind claims that Sefficiency increases *linearly* over time, or that selection is the *only* factor responsible for the origin and distribution of any particular trait. Nonetheless, the power of Darwinism lies in grasping simple, fundamental guiding principles, and there's no arguing that the crucial one is that "selection picks the best" (or rather the *better*) available trait relative to an environment. Hence, with these provisos, we can reasonably expect S-efficiency to show a positive trend—*if* the environment grants consistency and stability.

From what I've said so far, it may seem like biological traits always have Sfunctions and cultural traits always have D-functions and U-functions only. But this is not the case: any trait, biological or cultural, can have at any time both types of functions and efficiencies. For example, biological traits like abdominal muscles have a clear S-function of moving the torso; but if one intentionally works them out to obtain a sexy six-pack, they may be said to have an intentional D-function. As for cultural traits, countless examples can be found that have both D-functions (e.g., iPhones designed for calling and navigating the Internet; battle axes for fighting; Princess Diana's banged-bob hairstyle presumably for its desired look, etc.) and the S-functions by which they proliferate (respectively and hypothetically: signalling wealth/status; showing physical prowess; advertising trendiness).

I suggest that although non-technological cultural traits might often emerge through a selective process, and thus often have S-functions, their function is often of a different nature, obeys different dynamics, and is generally harder to understand (i.e. epistemically opaque) than technological traits' functions and efficiencies. One first suggestion is that non-tech traits may often have something more akin to S-function (not always, but more often); however,

due to the features of S-function described above, this might be more obscure to agents. Thus the problem may be not so much that efficiency is *missing* in the non-tech domain, but that "better performance" in relation to their S-function has completely different criteria, which are not captured by the standard examples about technological efficiency. As a result, when it comes to evaluating the criterion of increasing efficiency in cumulative culture, we should separate the discourse on technological cultural traits from that of non-technological cultural traits. In what follows, I explain why technological and nontechnological cultural traits are different with respect to function, and why the RA (or at least some versions of it) may be better suited to explain the evolution of cumulative culture in the tech domain.

Cultural traits (tech and non-tech) can have both D-functions and Sfunctions, but the crucial difference between tech traits and non-tech cultural traits is *where* they explicate their function. Efficiency is always relative to contexts or environments: technological traits mainly display their functions (Sand D-function) in the physical, material environment, while non-tech traits have functions (S- and D-functions) in social, symbolic, abstract environments. The fundamental difference between these environments is their respective stability and internal coherence. The material/physical/inanimate world is fairly consistent. On Earth, gravity makes things fall to the ground, day follows night, sharp objects penetrate surfaces better than blunt objects, water puts out fire, The physical systems tend towards entropy, etc. (human) symbolic/social/abstract world, by contrast, is highly inconsistent. On Earth, anything can become a symbol for anything, without there being any particular causal or physical connection between symbol and referent (arbitrariness); agents can have conflicting goals, interests, desires; they can lie or be truthful to themselves and others; they can be irrational and highly influenced by other agents. Because function and efficiency are relative notions that depend (among other things) on the nature of their environment, functions and efficiencies depend on the stable or unstable nature of their target environments.

Let's disentangle this further. The D-function of tech traits is generally much more epistemically transparent to agents than that of non-tech traits. I would go as far as maintaining that material D-function is *entirely* transparent to the designer of a tech trait. Certainly designers of toothbrushes or bottles have teeth-brushing and liquid-holding in mind when they design these objects. This does not mean that designers will always produce successful (efficient) objects or behaviours. Many a "Darwin Awards" winner may be listed in this category of "misguided designer", i.e. those designers that have a clear D-function in mind, but due to lacking knowledge and/or understanding of the environment in which the trait has to work, produce inefficient traits. (Wikipedia has a list of "Inventors killed by their own inventions" for a bounty of examples.) Nevertheless, provided designers have acquired sound knowledge of the environment and of a trait's interdependencies, the D-efficiency of tech traits will be high—and most importantly, will be aligned to the traits' S-efficiency (provided that they don't acquire social or symbolic meanings—in which case, the tech S-function could overwrite the D-function and smother its effects).

Notice that knowledge of the relevant environment is knowledge of the *material, physical*, mostly *inanimate* world for tech traits. Though complex, the physical world is relatively stable and predictable: unless gravity suddenly inverts its course, hammering a nail downwards will always be more efficient than hammering it upwards—and learning such a thing will be a permanent acquisition that doesn't need revision. The D-function of hammers may shift across slightly different tasks (driving nails, separating skirting boards, etc.), but these will always be underpinned by the same old physical rules. Hence, once correct knowledge is obtained, it will remain fairly robust over time. Importantly, correct background knowledge will ensure that D-efficiency is positively correlated to S-efficiency (which ultimately testifies/explains traits' success).

By contrast, non-tech cultural traits often appear to have an unclear D-function or no D-function at all, and consequently their efficiency appears

difficult to define and quantify. What is the D-efficiency of religious beliefs, kindergarten rhymes (Morin, 2015), futuristic poetry, ludicrous internet memes, and how can we measure it? How is a pencil skirt design "more efficient" than a bubble skirt? The D-efficiency seems hard to pin down but many (perhaps most?) non-tech cultural traits have S-efficiencies (or at least, so cultural evolution theories suggest). Recall that the "adaptive environment" of non-tech traits is the shifting ocean of human preferences, status badges, symbolic innuendos, social norms, fashion trends and so on. If we have to take the notion of selection seriously, the S-function of non-tech traits will have something to do with the reason why these traits originate, spread or fail to do so. Accordingly, there are many evolutionary hypotheses concerning religion's existence and proliferation in terms of its doing *something*. Thus there is nothing to stop one claiming that some religions are more S-efficient than others in having better achieved these "goals". Rituals, costly signals, norms and art may all have similar evolutionary explanations in terms of their S-functions.

So non-tech cultural traits may well *have* S-efficiency, but it remains problematic. First, there is the epistemic problem that it is often very hard to know what the S-function of a non-tech trait is (compared to a tech trait, for which such knowledge is often fairly transparent and aligned to D-function). Secondly, because non-tech traits are generally to be assessed for S-efficiency against a fluid social environment, their efficiency can wax and wane quickly and unpredictably. Non-technological cultural traits must "work" in a highly unpredictable environment: human psychological predispositions, which in turn are context-sensitive and frequency-dependent (due to norms and conformist biases). The "fundamental rules" of social and symbol-ridden environments do not afford such stability: what is fashionable today may be an utter no-no tomorrow (or the day after); overt goals may shadow covert ones.

These two problems with non-tech S-functions are obviously not unrelated to each other. The unpredictable fluidity of non-tech S-functions is part of the reason that we struggle, epistemically, to nail them down. The reason that the S- functions of non-tech traits are often quite epistemically opaque, and correlate poorly with their subjective D-function, is to be found in the nature of their adaptive environment: the social, abstract, symbolical domain of culture.

3.7.3. Efficiency, fitness and viability in technological traits

The previous two sections have shown that although non-tech cultural traits may have some form of function and efficiency, these behave in extremely different ways from their technological counterpart: ultimately, they cannot be considered equivalent notions. As a result, I argue that we should separate the discourse on tech and non-tech cultural domains, at least if we want to include *increase in efficiency* in the definition of "cumulative culture", as most RA adopters appear inclined to do.

I will finish this discussion of efficiency in technological traits with a final comment on its relation to fitness (to S-efficiency, indeed). As noted for the NK technological model, performance efficiency is explicitly compared to an informal notion of fitness. The idea is attractive, but as Lewens (2002) notices, not perfectly accurate.

[W]hatever property fitness is, it clearly has something to do with the expected reproductive output of some type of item. Hence, if we are to give fitness any kind of strong evolutionary reading, the term has to reflect, in some sense, the expectation that an artefact, or perhaps the plan, techniques or ideas that give rise to an artefact, will persist or proliferate in some community. (Lewens, 2002: 203)

The problem is that the efficiency of an artefact (how well it performs its dedicated job) is not always an obvious predictor of its diffusion, and we are often unable to claim that the "best artefact for some job is also the fittest in

that situation" (ibid.: 204). This doesn't seem to depend only on the dispositional or propensity-like characterisation of fitness in biology (i.e. the fact that fitness is always the *expected* number of offspring, not the actual: a "very fit" organism may be struck by lightning), but also on the peculiar nature of selective pressures faced by artefacts, which are at the "whims and fallibility of the agents who use or consume technology" (Ibid.: 204). These evidently represent a highly shifting, inconsistent, even contradictory adaptive environment—they are rife with symbolical and social features. An extremely efficient artefact or technique may fail to spread because its design or practice doesn't appeal to the current fashion, because for some unfathomable reason it is prohibited by law or religious dictates, because it fails to reach users due to a poor marketing strategy which didn't grant it enough visibility, or because it is too expensive compared to less efficient but cheaper versions.

I think that, other things being equal and with the appropriate degree of informality, it is not unreasonable to expect that the efficiency of technological objects correlates positively with their "fitness"—with their diffusion potential. In general, it seems true that the simpler the society (and the less intricate the network of norms and institutions, the degree of separation from manufacturing process, the potential for persuasion and communication, the association of technological artefacts to status or other socially relevant parameters), the more artefacts' "fitness" will be positively predicted by their efficiency. (Tellingly, in nonhuman societies, good and efficient "technological" tricks are readily preserved, even in the face of cognitive limitations in learning: think of the black rats peeling pine cones. Plainly, the procedure is copied because it *works*.)

In complex global markets of highly enculturated consumers, however, the relation between performance and diffusion of tech traits is often disrupted by their assuming overlaid *non-tech* functions. For example, while it seems correct to say that iPhones were originally invented for calling and navigating the Internet, much of their diffusion can now be explained by their being social status symbols (and their failure to proliferate further, perhaps, by their price).

Non-tech functions and efficiencies, in turn, have rather different dynamics and exceptionally turbulent "fitness".

In any case, for the present purposes of clarifying the RA relevant notion of cumulative culture, we may even dispense with the doubtful fitness/efficiency analogy altogether. What Tomasello's account seems to imply, in contrast to Jablonka's, is perhaps more akin to the biological concept of *viability*. Viability, unlike fitness which comes in degrees, is an all-or-nothing, discrete property: either an organism is viable or it is unviable. In technological traits, viability may be re-described as a threshold of *minimal* efficiency: a viable technological trait is one that is *minimally* efficient for a particular function (employable); an unviable trait is one whose efficiency has collapsed below the threshold to become unserviceable. (We can remove quite a lot of fancy bits from a car and still be able to drive it around somehow-perhaps more slowly and not on all terrains. It will be an inefficient car, but it can still perform its crucial function.) I think that the idea of complexity in technological traits is more often tied to viability than fitness, indicating the point at which deconstructing interdependencies "kills" the trait, or renders it useless. (Think again of the car parts in a pile or in a giant tumble dryer.)

3.8. Recapitulation and cognitive complexity

The discussion of Jablonka's and Tomasello's accounts of cumulative culture has forced a lengthy detour into the analysis of complexity and efficiency—which nearly eliminated their initial chronological characterisation as prerecapitulation criteria. In this section, I briefly describe the new rationale of recapitulation and its prospects for the identification of cumulative cultural phenomena.

This criterion essentially recommends that we label as "cumulative" only those cultural traits that go *beyond individual inventiveness*—that are too

complex for one individual to conceive, create or devise. To avoid this long periphrastic formulation, I name the criterion "recapitulation"—for each cumulative trait "recapitulates" the history of modifications that brought it beyond a single individual's understanding. It was first proposed in 2009 by Tennie, Tomasello and Call (Tennie, et al., 2009), subsequently becoming something of a mainstream trend (Dean et al. 2014; Dean et al. 2012). The very title of Tennie et al.'s paper ("Ratcheting up the ratchet") leaves little doubt as to the intellectual environment of the proposal. Recapitulation is in fact a revamped re-assertion of the decade-old RA and its Tomasellian notion of cumulative culture, with some refurbishing and even attractive restyling, but no substantial theoretical novelty.³⁸ As I have mentioned in the introduction, recapitulation represents a different (but cleverer) way of spelling out the old *accumulation of modifications* criterion.

One indisputable virtue of recapitulation is that it avoids the ambiguity caused by the level-neutrality in the underlying notion of culture—the ambiguity between Tomasello's and Jablonka's accounts of cumulative culture. It does so by exploiting the third notion of complexity identified in section 3.6.1, cognitive or epistemic complexity, and thus constraining complexity at the level of individual traits, learned and performed by individuals.

The centrality of epistemic complexity is evident in Tennie et al.'s formulation. According to these authors, the required evidence for a trait to qualify as cumulative culture is that it cannot possibly have been invented by a single individual—or, more technically, traits that are unambiguously outside the Zone of Latent Solutions (ZLS) of the individual. The ZLS represents that portion

³⁸ In this paper, formal but not conceptual refurbishment is also applied to the notions of social learning traditionally associated with Tomasello's work and the RA discussion, imitation and emulation. These are reformulated as process-oriented vs. product-oriented copying (as opposed to copy action morphology vs. copying outcomes). While this adds no theoretical difference, I think it hints even more clearly at the fact that the RA was formulated with gradualist considerations in mind, as I suggest with the gradualist interpretation in chapter 5.

of the space of behavioural solutions that is readily searchable via individual learning by any member of a species. Some latent solutions, i.e. non-cumulative traits, may be readily inventable and repeatable; others may be slightly more demanding, arising "rarely and only if all conditions are right" or thanks to "particularly gifted individuals" (ibid.: 2406). By contrast, cumulative cultural traits sit *outside* the ZLS, beyond the capacity of individuals' inventiveness.

With this formulation, Tennie et al. constrain single traits (at least in theory) to carry the observable marks of the accumulation process described by the RA. Their existing complexity "recapitulates" the history of modifications and/or collective contributions, and must be mastered by one individual. By shifting the attention from ontological complexity per se to the *cognitive* challenge posed by complex traits to individual capacities, recapitulation can conveniently apply to individual traits only—for individuals do not invent or perform collective repertoires (although humans may represent an exception to this). The reliance on epistemic rather than ontological complexity is perhaps the most interesting novelty of recapitulation.

Notice that cognitive complexity seems well aligned to both ontological interpretations of complexity, as I have already noted. However, I hypothesise that Simon's complexity poses a greater and more diverse cognitive challenge than Oswalt's. The sheer number of steps in a procedure, or the collocation of several rather than few foraging sites, will certainly place a toll on memory. However, if by accumulating modifications a trait accumulates complex causal and normative interdependencies, it will very likely become harder to *understand*, and as a consequence to learn and produce. If one has to build a handled hammer or copy the manufacturing procedure, one has to *understand* (as well as remember) that the head needs to be secured at the tip of the handle, rather than in the middle. Likewise, understanding that the grip needs to go on the handle rather than on the head of the hammer seems necessary for efficient and safe use of the tool.

A second alleged virtue is the empirical tractability of the criterion. In particular, recapitulation has been indicated as a promising tool for detecting cumulative culture in nonhuman animals in the most comprehensive review of comparative cumulative culture available to date. Dean et al. (2014) adopt recapitulation to evaluate claims of nonhuman cumulative culture, pointing explicitly to its practicality: "[F]or practical reasons, a useful yardstick is that the character should be beyond what a single individual could have invented alone" (ibid.: 5).

Interestingly, however, by adopting recapitulation these authors ultimately rule out some of the most promising examples of nonhuman cumulative culture in the wild, including various chimpanzees' techniques (nut cracking, ectoparasite manipulation and inspection, digging of wells in dry environments with the addition of leaf sponging, and more, cf. Boesch, 2003), and New Caledonian crows' stepped tools described by Hunt and Gray (2003). The examples are all discarded on grounds that it is highly unclear whether the trait is really outside the species' ZLS or could in fact have been devised by a single individual—evidence about the process of social acquisition and about the relation of the current artefact morphology to previous versions seems still required.³⁹

This conclusion, I believe, teaches an important lesson about recapitulation. Accumulation of modifications, especially in the cases of relatively simple behaviours and artefacts, leaves no clear-cut and unequivocal signature. We may try to salvage the criterion, as Dean and Laland do, by imposing the "burden of proof" standard, calling for painstaking demonstration that each modification has been introduced via social transmission rather than

³⁹ Some researchers of applied comparative cognition express open scepticism towards recapitulation. One complaint is that the ZLS criterion seems theoretically unsound: if a behavioural trait is *beyond* the possibility of invention of a single individual, how can one individual come up with it in a first place? (Ludwig Huber and research group, personal communication during a seminar at the Messerli Institute in Vienna).

individually produced. But if this is the way to go, it is unclear how recapitulation would make detection of cumulative culture more empirically tractable (it instead requires an onerous amount of fieldwork), or in what ways exactly it would be different from the good old Tomasellian criterion of *accumulation of modifications*.

The situation might be slightly different in experimental contexts. A few authors (Horner et al. 2006; Whiten & Mesoudi 2008) have attempted to implement recapitulation in experimental settings by presenting groups of children and chimpanzees with artificial fruits that required a step-wise procedure in order to be opened, with each step making the next step available. The idea is to check whether, having stumbled by individual exploration/chance on each procedural step separately (each procedural step being ideally within the individuals' ZLS), individuals are then able to retain each innovation and combine it in the correct opening sequence. Results have so far demonstrated an overwhelmingly good performance in human children, and only poor results in chimpanzees (Horner et al. 2006; Whiten et al. 2009), but more evidence is required to draw robust conclusions.

3.9. Conclusion

It is time to summarize the message from different accounts of cumulative culture. In section 3.5, I contrasted Jablonka's and Tomasello's accounts and I presented them as the two major competing alternatives. This was something of a narrative expedient to show as vividly as possible how things that equally go under the name of "cumulative culture" can be deeply diverse, and how this ambiguity is immanent in the literature. The two accounts can be plugged back into the RA Abacus scheme as they are—and as I will eventually do; however, combined with previously identified conceptual distinctions about culture, they logically generate *four*, not *two*, principal ways of thinking to cumulative culture.

We in fact need to take into account the *levels* factor that was identified as a possible source of ambiguity in section 3.2.

Recapping: the basic distinction we need to draw, at least in the context of RA explanations, is between accounts of cumulative culture based on quantitative complexity and accounts based on qualitative complexity—for brevity, *quantitative* vs. *qualitative* accounts of cumulative culture. Their crucial divergence lies in the differential ability to portray organically the dimension of efficiency and/or viability, which in turn, I think, suggests that we can properly apply qualitative accounts only to cumulative *technology*, rather than to *generalist cumulative culture* (i.e. including non-tech cultural phenomena). Thus the quantitative/qualitative distinction largely prefigures the tech/non-tech dichotomy. When this intersects with the level distinction, it produces four possible ways of thinking to cumulative culture:

(a) **Qualitative-individual cumulative culture**: identifies the qualitative increase in Simon's complexity of *individual-level cultural entities*.

The paradigmatic referent of this account is the *technological trait*—a skill, a procedure, and particularly an artefact.

What accumulates are (interdependent) modifications to the trait.

(b) **Qualitative-collective cumulative culture**: identifies the qualitative increase in Simon's complexity of *population-level cultural entities*.

The paradigmatic referent of this account is the *technological repertoire*. What accumulates are (interdependent) modifications to the repertoire.

(c) **Quantitative-individual cumulative culture**: identifies the quantitative increase in Oswalt's complexity of *individual-level cultural entities*.

The paradigmatic referent of this account is the *generalist cultural trait* (including non-tech ones). A particularly interesting case in this category are non-tech traits (symbolic, normative, ritual, linguistic, etc.).

What accumulates are (additive) modifications to the trait.

(d) **Quantitative-collective cumulative culture**: identifies the quantitative increase in Oswalt's complexity of *population-level cultural entities*.

The paradigmatic referent of this account is the *generalist cultural repertoire*, which includes tech and non-tech repertoires.

What accumulates are (additive) modifications to the repertoire.

This classification of cumulative culture notions is clearly not watertight nor exhaustive. In particular, one needs to notice that the level distinction is somewhat artificial, for the accumulation of modifications at the level of individual traits will sooner or later bud into accumulation dynamics in the repertoire, to the extent that some modifications produce novel traits (although I will not enter the complex topic of behavioural novelties here, see Brown 2014). Furthermore, I do not intend to claim that this four-pronged classification always denotes clear-cut, self-standing, systematically defended accounts of cumulative culture in the literature; nor perhaps that it exhausts all possible ways of defining cumulative culture. (For instance, there may be even more fine-grained distinctions that take into account other factors of ambiguity discussed in this chapter, such as the substrates of culture in section 3.2.)

Nonetheless, I think the classification offers a reasonable primary guide to disambiguate notions of cumulative culture, particularly when it comes to evaluating proposed explanations for their emergence—which is the approach we need in the RA context. Elements of these perspectives are fairly easily recognizable in actual RA-framed discussions of cumulative culture. In what follows, I will use the four-pronged classification to isolate a set of accounts of cumulative culture that have been taken as RA's explananda, and I briefly illustrate them via concrete examples.

The first possibility is the one that I have discussed under Tomasello's name and contrasted with Jablonka's account: a qualitative, individual-level
picture of cumulative culture. The paradigmatic target of this account and the connected RA explanandum, as noted before, is the single technological trait, most often an artefact, though the account unproblematically extends to behavioural procedures and skills. Examples may range from the now familiar hammer-like tool to more complex machines like pulleys or engines. But, as noted, we may also enlist technological procedures and skills, such as specific behavioural steps in a tuber-detoxifying procedure; or the exact sequence and topography of actions required to fix a stone tip to a shaft with adhesive technologies.

Typically, this account treats single tech traits as analogues of biological organisms—things with an overarching teleological organization whose parts need to be correctly organized to function and be viable. The "parts" of the Simon-complex system connected through interdependencies are sub-individual details of a single cultural trait, mostly physical parts or behavioural subcomponents in generative procedures (e.g. the actions in a behavioural recipe required to build an artefact, see Charbonneau 2015). It is also not uncommon for these technological traits to be conceived of as "cultural adaptations", with a prime example being the work of Boyd, Richerson and Heinrich (Boyd, Richerson & Heinrich 2013; Heinrich 2015). These features generally introduce a marked selectionist slant in RA-based explanations: the efficiency or adaptivity of a trait (which depends on its organization and ultimately on imitation in L1 in the RA line of reasoning) are part of the selection-based explanation for its current maintenance and distribution. The individual, qualitative, technological account of cumulative culture is perhaps the most common and (I think) interesting interpretation of this notion in the RA—a very narrow interpretation indeed. The problem, if there is one, is that the literature often fails to explicitly recognize and accept such narrowness.

This is one possibility; but the qualitative notion of accumulation and complexity may also be seen as applying at the *collective* level. This would deliver, I think, something akin to the idea of combinatorial culture that Balfour

was fond of. While the idea is often toyed with in the literature, I do not think that there are currently well-developed accounts of this kind. In addition, the available accounts that appear to move in this direction seemingly take *general* culture as their explanandum—or at least they do not operate a sharp distinction between tech and non-tech domains. A modern account of combinatorial culture may be exemplified by the "collective brain" hypothesis, recently advanced by Muthukrishna and Henrich (2016). In combinatorial culture, the Simon-complex system is the global cultural repertoire, and something very similar to interdependencies hold together traits and even groups of traits in complex, interdependent ways, generating functional wholes (e.g. institutions, enterprises, large-scale technology?).

While this account certainly contains elements of the qualitative-collective perspective I have indicated above, it exceeds the scope of my discussion in various ways. For one, the analysis of interdependencies undertaken here (as teleologically organized causal, physical relationships between parts) is clearly too narrow and simplified. Presumably, the teleologically organized framework will still hold, but the connections between high-level "parts" will be of a much more complex and varied nature, will include non-tech entities, and will be largely mediated by linguistic and other symbolic constructs. The "collective brain" hypothesis discussed by Muthukrishna and Henrich is an extremely thick, rich, and anthropocentric account of cumulative culture (see 3.1), for only in humans has the flow of cultural information reached a scope and integration such as to allow the meaningful recombination of inventions and repertoires' modifications at the global level. As such, I think it automatically diminishes the RA's explanatory potential; it patently calls for extra explanatory elements.

In general, I think that the RA is naturally challenged by collective accounts of cumulative culture (see the next chapter and the discussion of fidelity at the individual and collective level). But in the case of qualitative and combinatorial accounts, such as the "collective brain" hypothesis, there is the extra difficulty of fleshing out the nature of interdependencies at the higher level of cultural

repertoires. In conclusion, because of these important differences and because of the lack of clearly recognizable RA interpretations of rich combinatorial culture, I will tentatively include this account of cumulative culture in the final Abacus diagram, but I will not discuss it as part of a self-standing explanatory path.

(I cannot, however, resist adding a final speculation about efficiency and interdependencies in this account. As noted in the previous discussion, the kind of efficiency that cumulative technology theorists seem interested in is agent-based D-efficiency: it is this efficiency that is described as increasing in cumulative technology (habitually at the level of individual traits). D-efficiency clearly requires intentional agents. In this respect, Muthukrishna and Henrich's choice of the "collective *brain*" metaphor appears revealing: it seems that even repertoires, no matter how large, must be brought back (recapitulated?) to a unifying agential perspective in order to be treated as qualitatively increasing in complexity.)

Let us now move to quantitative accounts. Jablonka's account exemplifies the *quantitative* account at the *collective* level, generating as explanandum something akin to an "inflated" version of plain culture, i.e. a case in which cultural traits get preserved, but we don't see an increase in the complexity of individual traits. This will certainly represent an important "bead" in the concluding Abacus diagram. I have already given concrete examples of this account with macaques' "cumulative culture" on Koshima Island; nonetheless, one may see the quantitative-collective perspective as depicting cases of extreme quantitative accumulation in *human* repertoires of the general (including tech and non-tech) kind—which will clearly look very different from the Cercopithecinae's ones. In sections 4.2.1 and 4.3 of the next chapter, I describe Laland and Lewis' account of cumulative culture, which is modelled on the human paradigm, as another potential example of quantitative-*collective* account of cumulative culture (Lewis & Laland 2012). At this point, however, one may wonder whether the referents of Laland and Lewis' account on one hand, and of Muthukrishna and Henrich's on the other, are different in any significant sense. After all, both seem to capture complex, cumulative, ultimately general human repertoires. However, to the extent that an account of cumulative culture involves reference to its etiology (which appears nearly inevitable, since one has to specify what the diachronic process of "accumulation" means), it seems legitimate to discriminate between them. All the more so in RA contexts, where the etiology of cumulative culture is precisely what is at stake.

We have seen that quantitative accounts (at either level) do not advance any hypotheses about *how* parts are connected, and fail to provide a proper dimension for efficiency and qualitative directionality of cultural evolution. As a consequence, I argue, they are less explanatorily demanding and less explanatorily informative. My opinion is that they may be able to generally describe but not *explain* interesting cumulative dynamics (beyond the mere piling up of elements). The qualitative "collective brain" hypothesis described above may be inadequately tackled by the RA as it is; but it promises to recruit different instruments to *explain* how interdependencies work at the collective level. Quantitative *collective* accounts like Laland and Lewis' one, in contrast, risk dissolving the RA logic rather than resolving it or simply demanding integrative explanatory tools. I will discuss in the next chapter how such quantitative accounts tend to trivialize the content of L1 of the RA, and appear on closer inspection to be tautological lines of reasoning or little more than definitions of cultural processes.

Incidentally, quantitative accounts of cumulative culture tend to incorporate a less markedly selectionist perspective in the RA than qualitative ones. By missing a (D-)functional dimension of traits and repertoires, they appear amenable to the populational brand of evolutionary explanations (as in Mayr 1994), but not so much to selectionist one. If my association of quantitative perspectives and generalist, non-tech accounts of culture is on the

right track, this appears to nicely fit the tendency of attraction-based explanations to apply better to generalist, non-tech, linguistic and symbolmediated traits, and the corresponding tendency of selectionist ones to apply to technological adaptations.

Finally, there are quantitative-individual occurrences of cumulative culture. Laland and Lewis' account, in reality, may also be read as a quantitative*individual* example of cumulative culture—it depends on whether and when we are willing to consider a modified trait as a *new* trait. (The authors are ambiguous on this, in my opinion.) I think that a particularly interesting referent of this category (quantitative accumulation at the individual level of traits) is represented by *non-tech cultural traits*. Accordingly, in chapter 5 I will isolate a distinct version of the RA that takes as its main explanandum symbolical, ritual and social traits, i.e. things like social norms, dances, ritual implements, artistic creations and linguistic conventions. As we have already seen, these cultural traits, in contrast to technological ones, often have an arbitrary form and less clear functions.

CHAPTER 4. What is Fidelity?

4.1. Introduction

This chapter is dedicated to examining *high fidelity* ("fidelity" for short), as chapters 2 and 3 were devoted, respectively, to *imitation* and *cumulative culture*. Before diving into fidelity debates, however, I shall briefly chart out where we stand with respect to the overall picture.

In chapter 1 I argued that the RA can be stripped down to a minimal logical outline of three elements connected by two main theoretical links:

L1: connecting imitation and high fidelity

L2: connecting high fidelity and cumulative culture

often combined into a further link connecting imitation directly to cumulative culture, L3. This outline will come into play more frequently in these last two chapters, as the competing versions of the RA are ultimately outlined.

I further argued that within the boundaries of this simple logical structure, the RA is in fact not univocal, but rather ambiguous and divergent. This depends essentially on two factors: first, the *definitional instability* of the RA terms, i.e. the existence of multiple, tacitly fluctuating meanings of "imitation", "cumulative culture" and "fidelity"; second, the *explanatory gaps* left by the propensity to read theoretical links as necessary or probabilistic conditions (i.e. L1: <imitation is necessary/conducive to high fidelity transmission>; L2: <high fidelity transmission is necessary/conducive to the evolution of cumulative culture>). I have argued that in fact necessity and probability-based claims are merely *descriptive* of a status quo, and thus explanatorily defective: they can inform us, for example, *that* imitation is

conducive to high fidelity transmission; they can be used to debate *whether* or not high fidelity is necessary to cumulative culture; but they largely fail to walk the extra mile and clarify *why* and *how* imitation is conducive to high fidelity or high fidelity is necessary to cumulative culture (if this is the case).

In reality, asserting that L1 and L2 are *completely* bereft of explanatory details is a slight caricature of reality: from time to time authors have endeavoured to explain *why*, for example, imitation should be conducive to high fidelity transmission (some ideas, for instance, are briefly discussed in Acerbi, Tennie, & Nunn, 2011; Heyes, 1993) or *how* high fidelity is supposedly necessary to cumulative culture (examined in what follows). But, as will become evident in these last two chapters, these explanatory efforts appear for the most part out of focus, sparse and unsystematic; they are outshone by the "war of counterexamples", and the focus remains on the sheer existence (or absence) of necessary or probabilistic relations among the terms.

In sum, because both terms and theoretical links in the RA are considerably open to interpretation, one can obtain multiple versions of the argument by shifting among different meanings of the terms and different explanatory content of the links. The problem is that the existence of such diverse "logical paths" within the basic RA structure, and de facto of several different RAs, goes largely undetected: the RA is routinely presented and interpreted as a single, solid and relatively welldefined hypothesis (see Table 1 in chapter 1). I will ultimately come to extract and compare a few competing and essentially undetected RA versions in chapter 5, showing how they have all been actively maintained in the literature and how they have generated confusion and unnecessary disagreement.

Chapters 2 and 3 have made some progress towards this overarching goal by disambiguating among various relevant meanings of "imitation" and "cumulative culture". The current chapter undertakes the same exercise of disambiguation for the notion of "fidelity", and also embarks on the forward-looking task of resolving the explanatory gaps, which will eventually come to completion in chapter 5. The reason the present chapter has this double focus (and is consequently longer than

the previous two) is that fidelity is central to both L1 and L2, and its definitional volatility is very much influenced, I believe, by the theoretical role it is supposed to play in the two different RA claims. The chapter is hence organised as follows.

Section 4.2 exposes a prominent case of disagreement about L2 connecting fidelity and cumulative culture: one party claims that fidelity is necessary and crucial to cumulative culture; the other claims that it is largely insignificant and unnecessary. Conflicting opinions would be unremarkable, if not for the fact that they actually appear to be essentially *simultaneously correct*. Naturally, there's no deep logical puzzle behind this, but just another cases of definitional instability: the two conflicting claims use divergent meanings of "high fidelity", which I identify roughly as "action-matching" and "trait preservation" (respectively, fidelity qua AM or TP in the main text).

Section 4.3 takes a step back and clarifies that the meaning of "fidelity" *per se* is not in fact equivocal: fidelity is essentially a measure of degree of *similarity* (or in other words, it is errorless transmission, where the error that counts is mismatching—lack of similarity—rather than malfunctioning). The ambiguity is caused by applying the concept to cultural objects at different temporal and spatial levels (much like in Jablonka's and Tomasello's account of cumulative culture); thus the term "high fidelity" ends up denoting the two referents mentioned before: fidelity at the low level, or AM; and fidelity at a higher level, or TP.

The next problem I examine (in 4.3.2) is that many interpreters and consumers of the RA ostensibly fail to recognize such level-discrepancy and apply low-level fidelity in L1 (roughly, the previously identified AM) and high-level fidelity in L2 (roughly, the previously identified TP). However, these two notions of fidelity are genuinely divergent and need to be reconciled to make the argument sound.

Section 4.4 looks at existing proposals to "bridge" the level gap. However, by looking at this work, a tripartite rather than bipartite distinction of fidelity emerges. Fidelity appears to describe similarity of transmission: at the very low level of sub-components of an individual trait (fidelity-I, of which AM is an instance); at the intermediate level of the individual trait (fidelity-II, represented by TP); at the very high level of repertoires and trait distributions in the population (fidelity-III). This three-layered analysis of fidelity concludes the disambiguation goal of Chapter 4. As with AM/fidelity-I and TP/fidelity-II, fidelity-III also appears isolated from its predecessors: it needs more than just zooming in and zooming out to leap from one fidelity level to the other. Despite originating from the same definition (i.e. degree of similarity) applied at different levels, these notions of fidelity pick out different and incommensurable phenomena.

The rest of the chapter is devoted to shedding further light on the theoretical role of fidelity in the RA. To do so, I bring fidelity back to its native conceptual environment of biological evolution. The last sections deal with the idea that fidelity is equivalent to cultural inheritance (4.5), and that the requirement for extremely high fidelity in cultural transmission (down to the level of trait's sub-components, fidelity I) comes from likening cultural inheritance to the genotype-phenotype model of inheritance, or better yet to the model of so-called "Weismannian" inheritance—the inheritance of a phenotypical template (4.6 and 4.7).

The Weismannian model of inheritance, however, introduces a conceptual shift in the object of transmission: if what is transmitted are the "instructions" to build the cultural trait's "phenotype", then we need to establish an appropriate framework of reference in the cultural domain. What counts as the instructions and what as the phenotype? I think that the answer to these questions, as well as the overall appropriateness of the Weismannian model for cultural inheritance, is far from clear. Some considerations about the nature of cultural variation may support this picture, but other remain dubious (4.8). Be that as it may, one of the most common interpretations of the RA, which I will label "gradualist interpretation", appears to adopt exactly this conceptual model

and to derive its requirement from high fidelity transmission of individual traits sub-components from it.

4.2. Competing notions of fidelity

Despite its theoretical significance, fidelity has received remarkably scant press, both in the biological and cultural domains. As far as the RA is concerned, while virtually everyone appreciates that a modicum of terminological confusion surrounds the notions of cumulative culture and imitation, what we mean by "fidelity" has gone largely unquestioned. Rather than analysing and comparing competing accounts, as I've done for the two previous terms, it will be a matter of carving them out from more or less implicit, scattered cues.

In this section, I examine the position of two prominent contributors to the cultural evolution debate with respect to connection between high fidelity and cumulative culture (i.e. about L2 of the RA). On one hand, there will be the views of a group of scholars that are generally identified as "cultural epidemiologists"; their work has its principal root in Dan Sperber (1996) and has recently been defended by Oliver Morin (2015). Cultural epidemiologists maintain that cumulative culture phenomena occur constantly and effortlessly (particularly in our species) without high transmission fidelity, and therefore that high fidelity is irrelevant and unnecessary. On the other hand, I will discuss an important piece of work from Lewis and Laland, which may be seen as broadly interpreting the views of all those who sympathise for RA-style ideas. Lewis and Laland are so resolutely convinced of the importance of high fidelity transmission to cumulative culture that they literally entitle their paper "Transmission fidelity is the key to the build-up of cumulative culture" (Lewis & Laland 2012). Their position is thus on the furthest end of the spectrum of RA claims about the high fidelity/cumulative culture connection, considering it as vitally necessary. Although this is a crude précis of the relevant opinions, which are part of much richer, subtle and extensive analyses, their contraposition on the matter is genuine. I will also show that once these arguments are subjected to closer scrutiny, both positions appear fundamentally on the right track. This apparent paradox, I claim, is due again to an ambiguity in the notion of fidelity. But let us take a closer look at these competing views.

4.2.1. Sperber and action matching

The most radical claim against fidelity comes from the cultural epidemiologists' group, with which Morin openly self-identifies. His work will be my main concern here because it is the most recent and the most directly and specifically critical of RA ideas and particularly of the imitation hypothesis. The suspicion about the constitutive inaccuracy of cultural transmission had already emerged towards the end of the 1990s, at the hands of three important figures in the field of cognitive anthropology (Boyer, 1994, 1998; Sperber, 1996; Atran, 2001, 2002). These authors criticised what they called the "Neo-Darwinian" approach to culture,⁴⁰ which they took to perpetrate a "serious distortion of the relevant facts" (Sperber 1996:118) by considering cultural transmission adequately modelled on the biological concepts of high fidelity replication and reproduction.

Cognitive anthropologists observed that cultural transmission, by contrast, is better characterised as a transformative process, whereby mental representations are not "replicated" or "reproduced", but rather

⁴⁰⁴⁰ Which in their view includes the whole field of memetics (Dawkins, 1976; Dennett, 1993); socalled "dual inheritance" models developed by Boyd and Richerson (Boyd, 1985); Cavalli-Sforza and Feldman's quantitative approach (Cavalli-Sforza & Feldman, 1981) and Lumsden and Wilson's psychology-driven picture of cultural evolution (Lumsden & Wilson, 1981). Later, Sperber explicitly acknowledges that his early critiques of Boyd and Richerson work were misdirected (Claidière & Sperber 2007).

"reconstructed" or "re-produced" over each transmission event, through an inferential process influenced by what Sperber came to call "cognitive attractors". As a result, they claimed, cultural transmission is typically inaccurate, or low fidelity. More specifically, Sperber's attack against fidelity may be articulated into three main argumentative lines.

First, he proposes what I will call the "argument from translation". This is based on the observation that cultural transmission always involves an alternation of mental/private and material/public representations. For example, the tale of Little Red Riding Hood that I hold in my head is a *private* (essentially, mental) representation; if I want to tell you the story of the little brave girl who falls victim to the big bad wolf's schemes (and appetites), I must necessarily translate the private representation into a *public* one: this may be an oral bedtime story, or an illustrated book, or a shadow puppetry show—any medium that puts the story "out there" for transmission. As I have already noted in chapter 3, Sperber and his associates take *mental* representations to be the fundamental substance of culture—the representations that matter (Sperber 1996: 24); but no mental representation can be passed on unless translated into a visible, audible or otherwise sensible public form-this is obligatory and happens both ways at each transmission episode, inasmuch as the receiver has to re-encode the public representation "up" into his mind and then "down" again into a further public manifestation (if he wants to pass it on). This argument may be read as pointing to the lack of a common substrate/medium for mental and public representations, which is likely to result in loss of information, much like a translation between different languages cannot ensure perfect smoothness (the "lost in translation" conundrum than any foreign speaker experiences when she ventures into excessively fine-grained clarifications). However, it is generally and perhaps more compellingly a specific case of what Morin calls the "wear and tear problem", i.e. the empirical fact that the longer a transmission chain, the more copying errors tend to accumulate along it (unless correction mechanisms intervene). Information decay and

erosion over transmission episodes is a pervasive theme that is often effectively illustrated with the example of Chinese Whispers (cf. Ridley 2000; Rendell et al. 2010; Enquist et al. 2010). So the constant translation required between public and private, by multiplying transmission events, may make transmission less accurate over time compared to a direct transferral method (e.g. privateprivate).

Second is the argument for "cognitive alteration". The nature of cultural transmission, claims Sperber, is interpretative, transformative, creative, and crucially hinges on the mental environment of the agent. Internal representations (the crucial ones in his view) are subjected to the influence of other cognitive processes and pre-existing content in the mind of the agents, which inevitably modifies them. (For example, empirical evidence shows that asking siblings to report shared past events results in strikingly different stories, French, Sutherland, and Garry 2006; more trivially, the phenomenon may be familiar to any couple who have tried collectively to recall the trigger of an argument: memories are often frustratingly divergent.) Heyes 1993 makes a similar remark about lack of insulation from other mental content potentially leading to lower fidelity in social learning. Somewhat in connection to this argument, Morin frequently notes that we are not "wired for repetition", or in other words that we are not imitators by default. We fortunately feel no compulsion to copy each gesture our teachers present us with: we do not reproduce every pointing gesture when given directions-that would disastrously hamper prospects for effective communication and make our life a misery (cf. Morin 2015: 73). In other words, our *motivation* (a cognitive, internal alteration factor) is not always that of ensuring mimetic, conservative transmission of information: communication is not necessarily conservative, and it seems to constitute a sizable share of our daily informational transactions. (As we have already seen, this may also be seen as part of the flop problem faced by social learning, see chapter 2.)

Third is the "multiple sources" argument, which is often considered the crux of the claim against fidelity. Culturally transmitted ideas and representations, according to Sperber, are never *copies* of their predecessors; the process of transmission cannot generally be characterised as replication (see also Sterelny 2001 and 2017 on why this is supposedly important). In genes, replication constrains the ancestry relationship to a significant extent: replication is always (or nearly always) one-to-one or one-to-many, but never many-to-one. Gene replication is a duplication process in which a given unit serves as template for other copies to be made, and each gene copy can be retraced back to its singular predecessor (i.e. we can trace virtually univocal *lineages* of genes' replicas). By contrast, ideas (and "memes", which were among the original targets of critique) behave in a different way. The idea of the Eiffel Tower or of Bill Clinton (Sperber's example) that I have in my head is almost certainly the result of a number of parent-representations, i.e. my preceding encounters with representations of these real-world objects. During learning, I keep enriching my representational tokens through multiple and successive bouts of exposure, often coming from disparate sources. To the extent that the parent-representations are diverse, Sperber argues, their blending will dilute and corrupt fidelity of transmission. Thus, importantly, we cannot use the biological model of replication to describe representation transmission.

The case against fidelity may enlist a further argument, which is, however, much narrower in scope and thus weaker, for it can be applied only to cultural transmission that happens through linguistic means; I will call it the "argument from symbols". Atran observes that when a large sample of students was asked to write down (i.e. transmit) the meaning of Mao's *Little Red Book*'s phrase "Let a thousand flowers bloom", the results were (unsurprisingly) highly inconsistent (Atran 2001: 365); presumably this is due to the symbolical properties of language, which leaves ample room for metaphoric interpretations.

Collectively, the arguments from translation, from cognitive alteration and from multiple sources of representations (and perhaps the one from symbols)

demonstrate—according to Sperber and the cultural epidemiologists—that cultural transmission is on average low fidelity. (There are exceptions, they grant: but they are *exceptions*; and even some of these exemplary exceptions—amanuenses monks' work, for example—may not be clearly replicative, see Sperber 1996, chapter 5). The same claim about the customary low fidelity of cultural transmission is manifest in the work of attraction theorists, who are basically the most recent version of cultural epidemiologists, a rebranding. I will not get into in-depth examination of these arguments here, although I will discuss some of their points in greater detail further on; some of these arguments are subject to fairly obvious objections. (In particular, the claim about blending inheritance can be rejected by maintaining that averaging the common elements increases rather than decreases fidelity.) However, they are certainly right that many details get lost in transmission. The main goal is instead to understand which notion of fidelity these authors assume in putting forward their attack.

How is fidelity understood in this work? There is little explicit treatment of the notion, but the usage of synonymous expressions such as "accurate", "detailed", and "fine-grained transmission" suggests that, broadly speaking, the fidelity that cultural epidemiologists want to employ is something like the finegrained, high-resolution reproduction of the details of single behaviours' or cultural traits' occurrences. For example, Morin claims: "Imitators must reproduce at fine-grained level (fidelity) an action ... only faithful reproduction should be taken into account in our definitions of imitation" (Morin 2015: 57). In some parts of Morin's work and elsewhere in the literature, such faithfully reproduced "details" are of rather general and unspecified nature: presumably, they will be details about plot or wording if we are talking of the social transmission of a tale; details about material, shape and fixtures if we are talking of an artefact; details about the precise mirroring of a model's actions if we are talking of a particular behaviour (such as the familiar flint-knapping technique demonstrated to Abu by Ina); and so forth. Notice that in all cases, the finegrained features that get transmitted (or fail to do so) are always details of *single* cultural occurrences, confined to *local*, spatially and temporally limited circumstances; even more, they appear to refer to the transmission of *sub-components* making up cultural traits.

An even narrower representation of fidelity, however, may be described as "action-matching". Abu and Ina's case comes back to us here: Abu's imitation of Ina's manufacturing procedure will be called "high fidelity" whenever Abu's actions match precisely those performed by Ina: holding the stone with a strong grip, lifting the left arm for striking at such angle, rotating the torso to maximise impact, etc. Incidentally, it is exactly this very narrow type of fidelity as action-matching that seems to be the main target of Morin's critique. When criticizing the "imitation hypothesis" (i.e. another avatar of the RA), in fact, Morin describes the slavish, fastidiously pedantic property ascribed to imitation by RA advocates specifically as the reproduction of visuo-motor features such as the model's gestures, actions, action plans, etc. (cf. Morin 2015 : 57, 69, 73).

In reality, the prominence of action-matching as a way of conceptualizing high fidelity in Morin's work is far from surprising: it may in fact be the most widespread way of thinking about fidelity, especially in connection to social learning, and paradigmatically to imitation. This may depend largely on the ubiquitous adoption of the so-called two-action method to test nonhuman animals for imitative abilities discussed in chapter 2. Two-action method experiments have historically had a profound impact on the conceptualisation of imitation in the RA, promoting action-matching as the fundamental signature of imitation. The concoction of imitation, fidelity and human uniqueness conjured up by the RA and supposedly tested via two-action methodologies has created an inextricable association between these two elements (imitation and actionmatching). The connections seem to be as follows: (i) we test animals through the two-action method to dispel the human uniqueness claim; (ii) experiments of such design are considered passed, and imitation demonstrated, when subjects can match action with the model as a result of observation, so imitation gets robustly associated to action-matching (sometimes to the point of making action-matching central to its definition, as in all contrastive accounts of imitation vs. emulation—which are by far the most popular, especially in discussing RA hypotheses cf. Caldwell & Millen 2008a; Wasielewski 2014); (iii) according to the RA, imitation is the most high fidelity mechanism; therefore (iv) action-matching must represent the property of fidelity in imitation. Although this line of reasoning is sketched in very simplistic terms here, and certainly doesn't represent well-informed opinions in the field of cultural evolution, I suspect that inattentive readers may have derived their understanding of high fidelity as action-matching from a very similar chain of thoughts.

As a consequence of the putatively established standard low fidelity of cultural transmission, cultural epidemiologists deny that high fidelity can be responsible or necessary for the phenomenon of cumulative adaptation and cumulative culture more generally. They claim instead that the important factor is what has become known as "attraction" (Sperber 1996): cultural representations appear stable and recognizable because they tend to converge towards common parts of the entire space of possible representations ("attractors" as Sperber calls them—a terminology that appears to have won out over the similar notions of "cognitive elicitors" in Atran, and "triggered representations" in Boyer). Attraction tends to produce uniform representations because we all share broad inferential patterns and cognitive penchants: for example, we find geometrical shapes easier to retain and reproduce than random doodles; we prefer direct-gaze representations of faces over averting/covert gaze ones; the convergence on the use of the same term "red" to denote a particular spectrum of colours is due to relevantly similar physiology of perception in humans (Morin 2013, 2015; Claidière, Scott-Phillips, and Sperber 2014). We may re-construct information in mostly idiosyncratic ways at each episode of transmission and communication; but over time, our common mechanisms for filtering, transforming, neglecting and selecting will end up

highlighting and preserving the same core elements of representations, thus determining the shape of traditions.

The theory of attraction has encountered some resistance (StereIny 2001), but it has also recently undergone significant extension and refinement (Claidière, Scott-Phillips, & Sperber 2014; Acerbi & Mesoudi 2015), making it an increasingly important part of the debate on cultural evolution.⁴¹ It is worth noticing that this theory assigns a much less significant explanatory role to selection compared to other evolutionary explanations of culture: recently, Sperber has claimed that selection should be seen at best as a special case of attraction, relevant only in a subset of circumstances, while attraction represents the pervasive force guiding cultural evolution—if not cultural evolution altogether (Claidière, Scott-Phillips, & Sperber 2014; Claidière & Sperber 2007). Cultural epidemiology is still considered an evolutionary-based explanation of culture, both by proponents and critics, to the extent that it advocates the use of populational models to understand the distributions of cultural representations (Sterelny 2017): but the role of selection in the general picture is significantly diminished, and cultural "adaptations" are here explained by the gravitation of representation towards common points of attraction rather

⁴¹ For example, cultural epidemiologists now distinguish between attractors, factors of attraction and attraction. Attractors are points in the space of possible cultural variants where representations tend to converge; factors of attraction are the particular cognitive, psychological, physiological, historical or ecological structures or biases that make representations converge towards these points; attraction is the overall process describing the dynamic of cultural variants (Claidière, Scott-Phillips, & Sperber 2014; Acerbi & Mesoudi 2015). This finer articulation is intended to address the doubts expressed regarding the explanatory power of attractors (Sterelny, 2001a) when conceived simply as cognitive biases. While Sperber holds that the notion of attraction is substantially unchanged from its original proposal, it is fair to say that the present version has been developed, making it more obvious that we should not limit our understanding of attraction to a transformative process due *only* to *cognitive* factors—but that other sources of channelling should be taken into account.

than by a stepwise process of selective retention of the fittest variants, intergenerationally transmitted with high fidelity.

4.2.2. Laland and trait preservation

Let us now turn to the pro-fidelity paper by Lewis and Laland (2012). This work is explicitly directed at testing the RA and more generally the fidelity/cumulative culture connection (L2), similarly advocated by the likes of Galef, Heyes, Whiten and notably Dawkins (ibid.: 2172). The study specifically aims at providing a formal assessment of the link between fidelity and cumulative culture, for, despite numerous verbal endorsements, the authors maintain that "there has been no formal theoretical investigation of whether the hypothesised link between the two is correct" (ibid.)—one of the few explicit recognitions of the extant explanatory gap in L2 of the RA. The paper has enjoyed reasonable popularity, being among the first to apply a modelling approach to the question of social transmission's effects on cultural stability, maintenance and establishment—a direction that had long been perceived as broadly missing in the field (in contrast to more popular mathematical work on trait diffusion, cf. Enquist, Ghirlanda, & Eriksson 2011; Enquist et al. 2010).

As it is evident from their paper's title, the authors find the connection between fidelity and cumulativeness amply supported by their models; aside from the results, however, what is directly relevant to my argument is the fact that by having to include high fidelity among the models' parameters, the authors are forced to provide a very specific figure and definition of the notion.

Before I embark on an analysis of Lewis and Laland's account of fidelity, two clarifications need to be made. First, their paper draws heavily on previous work by Enquist and co-workers (Enquist et al. 2010), claiming to adopt *in toto* the results obtained about fidelity in this work, as well as the specific choice of formalisation of the notion (Lewis & Laland 2012: 2173); therefore, the discussion of fidelity reported here will be based on information from the two conjoint papers. Secondly, despite Lewis & Laland's explicitly purporting to provide evidence in favour of the RA, nothing in their work makes an explicit connection to imitation—nor to any other special mechanisms of social transmission. The same may be said of Enquist et al.'s paper (which addresses the RA even more ambitiously):

Tomasello ... argues that this transmission fidelity helps to explain the existence of cumulative culture (or "ratcheting") in humans but not in other animals. However, hitherto it has not been apparent why high fidelity should necessarily favour ratcheting (Laland et al. 2009). Our analysis potentially plugs a gap in this argument. (Enquist et al. 2010: 1361).

Given the absence in either paper of any explicit attempt to connect the fidelity variable to imitation (or to any other specific theorisation of social learning mechanisms in the relevant literature), I think that the results of these works should actually be considered pertinent only to L2 of the RA. This is, however, *not* what the authors suggest: in various places they claim to tackle the RA as a whole (i.e. L1 and L2 in connection).⁴²

Very briefly and for clarity, the basic description of the models in the words of the Lewis and Laland is as follows (most of the data is provided out-text in supplementary electronic material):

As a starting point, we assume that there are a fixed number of traits that can appear within a cultural group through novel inventions, independently

⁴² Notice also that they unproblematically incorporate and purport to substantiate through their analyses *attribution* claims about cumulative culture, imitation and other transmission mechanisms such as teaching and language-mediated learning in the argument, as customarily done in the literature; while as established in Chapter 1, attribution claims should be considered independent from the RA logic.

of any other traits within the culture. We call these novel inventions "cultural seed-traits" after [Enquist 2011] and for simplicity assume that there are up to 10 of these in a population. ... To begin, the cultural group is initialised with two cultural seed traits drawn at random. The next event that takes place can be one of four options. (Lewis & Laland 2012: 2173)

The four options are as follows: the addition of a new seed-trait drawn again at random from the external set; the *combination* of two cultural traits already present in the group, producing a new trait; the *modification* of one of the traits present in the group, also producing a new trait (a "variant"); the loss of one of the traits present in the group. Each of these events is assigned a certain probability (respectively: ρ_1 , ρ_2 , ρ_3 and ρ_4). At each event the culture of the group is updated (augmented with a new trait, or a trait removed) and the next event takes place. The population is taken through a total of 5000 events. Two sets of models are run. In one set of analyses—less computationally demanding—the parameters (modification, combination, addition and loss) are considered as dependent (each one's probability affecting the others', ρ_1 + $\rho_2 + \rho_3 + \rho_{4=}$ 1); while in another set—more computationally demanding and largely confined to the supplementary material—rates vary independently (ibid.: 2174). The results are understood as convergent across the two sets of analyses, and show that loss rate is by far "the most important factor in determining the cumulative nature of culture" (ibid.: 2176).

To understand where fidelity comes into play, we need to look at the "trait loss" parameter and relate it back to previous work by Enquist. Lewis and Laland assume that "a positive exponential relationship between transmission fidelity and longevity of cultural traits" was established by Enquist et al. (2010), and therefore use trait loss as a "proxy for fidelity" (Lewis & Laland 2012: 2173; 2175). From this claim, we can already conclude that fidelity is clearly understood as trait longevity or trait preservation in this work (being therefore in an inverse relationship with trait loss). But the inference becomes even starker by taking into consideration the original Enquist et al.'s modelling.

Looking back at Enquist' work, we find that fidelity has been defined as p, i.e. the probability that an individual acquires an observed trait over one social learning event (Enquist et al. 2010: 1356); p is always assumed to be <1, under the (reasonable) assumption that no social learning event and no transmission in general is perfect. In addition, it is important to note that they consider cultural traits as discrete, entire entities that can be viewed as either present or absent in an individual. For example, one individual could be described as either having or not having the trait "being able to drive a car" or "being able to fish for termites" or "being Buddhist" or "speaking French"; as a consequence, one can quantify unambiguously the number of French speakers or Buddhists in a population (Enquist et al. 2010: 1354). It also follows from this characterisation that traits are not, in these models, decomposable into sub-components: one individual cannot be represented as acquiring (say) the Buddhist belief about Nirvana but not the Buddhist precept of abstinence from intoxicating substances; at most, he might have higher or lower probabilities of acquiring the trait as an indivisible entity.

There is a good deal of confusion in the paper, because the authors refer in various places to the p parameter alternatively as the "fidelity", "success" or "efficiency" of social learning; at the same time, they use these terms to indicate the "probability that an individual has picked up the trait", presumably before becoming a "cultural parent" in its own turn (before being allowed to re-transmit the trait in the models).

Either way, fidelity (p) and the "probability that an individual has picked up a trait", which I may call P, are strongly correlated. The latter is in fact defined as:

 $P = (1 - (1 - p)^n)q$

with q being the probability that a cultural parent possesses the trait, and n the social learning trials with the same cultural parent.⁴³ In other words, if an individual selects a "good" cultural parent (one with high q), and can learn with high fidelity (high p), the probability that she will pick up the trait (P) increases over repeated learning bouts. Enquist et al. carry out their analysis across various sets of conditions (e.g. random vs. clue-guided selection of parent and therefore random or selected q; fixed vs. changing values of p for each individual; single vs. multiple cultural parent makes no culture", i.e. that learning from more than one model is crucial for the preservation of traits and the emergence of cultural phenomena. (Notice that this result may be seen in contrast with Sperber's "multiple sources" argument against fidelity.)⁴⁴

Although the formal definition of "fidelity" has a merely instrumental and tangential value in Enquist's work, as opposed to its centrality in Lewis & Laland's, it nonetheless allows one to see very clearly how fidelity is understood and conceptualised.

By being defined as "the probability of acquiring a trait", fidelity appears unambiguously defined as longevity or preservation of traits. The distinction between p (the probability that an individual acquires an observed trait over one social learning event) and P (the probability that an individual has picked up the trait) doesn't introduce any particular difficulty: both p and P are measures of

⁴³ This equation describes uniparental transmission; for multi-parental transmission, the relevant equation is: 1-(1-pq)ⁿ.

⁴⁴ They claim a number of other interesting results, all of which nicely fit empirical evidence and independent arguments made by other researchers. First, large populations provide a much better environment to the establishment of culture—the reason for this in the model appears to be that larger populations provide naturally more opportunities for learning trials. Second, fidelity seems to work with a threshold effect, especially in large populations. A small increase in transmission fidelity is necessary and sufficient to push the population's probability to possess the trait quickly above 90%, and therefore to possess culture. Third, in small populations innovations are likely to get lost immediately.

trait preservation at different grains of temporal analysis (this is particularly clear if you consider that if only one opportunity of interaction with the cultural parent is given, p entirely coincides with P); what is important is that fidelity measures concern the probability of preservation of an *entire* trait (as a discrete entity), rather than of sub-components of the traits like sub-actions or details of a cultural trait, like in action-matching and more generally the notion of fidelity addressed by Sperber and cultural attraction theorists.

I would also add that, given such definition of "fidelity", Lewis and Laland's results about fidelity and cultural accumulation are scarcely exciting (while Enquist et al.'s results about the interactions of population size, number of learning trials and fidelity thresholds are more genuinely interesting): if I keep having high probabilities of acquiring a trait over repeated trials, it seems no wonder that I will eventually acquire the trait, all other things being equal, and that I will eventually be in the condition to pass it on to someone else. Notice that the argument has striking parallels with what I have called "Jablonka's quantitative accumulation" in chapter 3: if traits get preserved in time rather than being lost, and other traits are gradually introduced, the overall number of traits will accumulate and lead to a larger cultural repertoire.

4.3. Levels matter: levels discrepancy in fidelity and problems for the Ratchet Argument

4.3.1. Fidelity as similarity: action matching and trait preservation

In the previous section, I have shown how Sperber/Morin and Laland/Lewis/Enquist positions on the fidelity/accumulation connection (L2) sit at opposite ends of a spectrum: fidelity is seen on one hand as crucial and necessary to the process of cultural accumulation, on the other as irrelevant and

unnecessary. It should be also apparent, however, that the contenders hold rather different conceptions on what fidelity is. Laland and Lewis think of it as the probability that offspring resemble their cultural parents with respect to the possession of specific, discrete traits, and refer explicitly to the positive correlation between fidelity and longevity of traits. For this, they seem to conceive of fidelity as something like trait preservation (TP henceforth). On the other hand, Morin, Sperber and more generally the cultural epidemiologists (or cultural attraction theorists) appear to think of fidelity as something like the reproduction of *details* of individual traits, and often even more specifically as the reproduction of details of motor behaviour in a social learning episode. For brevity, let us label this interpretation fidelity qua "action-matching", or AM henceforth.

These two notions appear fundamentally incongruous, or at least their connection is prima facie unobvious. At a closer look, however, one may notice that they derive from applying a uniform concept at different levels of analysis. In fidelity qua TP, what matters is that the temporally downstream individual (the cultural offspring) *resembles* its parent with respect to the focal trait. In fidelity qua AM, by contrast, the notion expresses the degree to which the sub-actions or behavioural components of the observer *resemble* the sub-actions of the model.

What I suggest is that the underlying meaning of "fidelity" is in fact not so divergent or inconsistent across TP and AM: "fidelity" means, essentially, *similarity*, or better degree of similarity between two entities—but the concept is applied to cultural entities at different levels of analysis, resulting in two divergent referents. In AM (or in other detail-oriented, non-visuomotor centred accounts of fidelity) the elements compared for similarity are *low-level* details of cultural entities: subcomponents of individual, local, temporally and spatially limited behaviours—most often individual episodes of social learning. If Ina holds the striking stone in the left hand with a strong grip, and Abu *similarly*

copied with high fidelity. Conversely, if after observing Ina, Abu holds the stone (say) with a precision grip, or with his right hand, or uses a strong grip on the anvil stone, Abu will have copied with low (or lower) fidelity.

In TP, the elements compared for similarity are *entire* traits, intended as discrete entities that individuals may or may not possess; or even trait distributions among the members of a population. If Linda is a piano player and teaches her daughter how to play piano, her daughter will have a certain probability of becoming *similar* to Linda with respect to piano-playing. When the probability of this similarity is high, transmission is said to be high fidelity; conversely, when the probability is low, fidelity will be considered low. The same happens for the group level: if a population has a high proportion of taxi drivers, and taxi-driving can only be learned socially, high fidelity transmission will produce a population with a *similar* proportion of taxi-drivers; but the parental generation of taxi drivers will be a bad predictor of the offspring generation of taxi driver if fidelity is low, because the proportion would be *dissimilar*. Notice that the acquisition of traits such as the one described by Lewis, Laland and Enquist will typically occur diachronically, i.e. over several bouts of social learning (this is indeed an in-built assumption of their models): one doesn't typically become a taxi driver, a piano player or a French speaker by fiat—the acquisition of the trait is a lengthy and often costly iterative process.

Interpreting fidelity as degree of similarity seems the correct way to go—in fact, other independent lines of evidence suggest that fidelity should be conceptualised as similarity (see biological argument in 4.5). This will in turn pose other issues, but we need not worry about these now; I will pick up again the topic of fidelity as similarity in the next section. Let us now look at the explanatory role of the level-specific notions of fidelity in the RA.

Not too surprisingly, we may observe an intuitive "good fit" of each of the isolated meanings of fidelity as AM and as TP with the RA claims L1 and L2. It is immediately clear that the first meaning, fidelity as AM, is quite useful to explain L1: as remarked, one of the most common reasons imitation is considered the

only or most faithful social learning mechanism is precisely that, by sourcing information directly from the morphology of behaviour, it has more probability of resulting in action-matching (i.e. in fidelity at the low level). Thus conceiving of fidelity as AM validates L1: imitation is particularly or uniquely conducive to fidelity qua AM because it is better placed than other social learning mechanisms in producing action-matching.

Likewise, fidelity as TP appears obviously suited to L2: as seen in chapter 3, any complex or cumulative process needs preservation; if things are not preserved, or fail to be transmitted, accumulation simply cannot occur (remember the hoarder). While TP validates L2 most obviously under the interpretation of cumulative culture as quantitative accumulation (roughly, Jablonka's interpretation), notice that TP is in fact the necessary precondition for *any* cultural phenomenon, be it simple, quantitatively, and even qualitatively cumulative. The relation between TP and simple culture may be conceived of as one of identity, inasmuch as high probability of trait preservation can be equated to trait transmission. In turn (as noticed in chapter 3) quantitative cumulative culture is hardly distinguishable from simple culture (recall again Jablonka): we need postulate only the regular addition of innovations and their preservation to see the size of the population's repertoire increase (which, incidentally, is what makes Laland's conclusions scarcely exciting). But we should remember that the same dynamic of preservation must occur also in support of qualitative accounts of cumulative culture. TP may not be sufficient to explain qualitative cumulative culture and its interdependency-based complexity, but inasmuch as it defines simple culture and cultural transmission, it must be considered a necessary feature (see Ch. 5 on quantitative interpretations for more on this). Thus conceiving of fidelity as TP offers a direct confirmation of L2 in at least one important sense: TP-fidelity is necessary for cumulative culture because without preservation there is no (quantitative) accumulation. (I will talk about qualitative cumulative culture in a moment.)

This is all well and good. TP and AM may be two individually legitimate ways of disambiguating the notion of fidelity, and may even be taken to independently explain something about imitation or cumulative culture, respectively. But the original issue was to understand fidelity *in the RA*; and the RA results from the *combination* of L1 and L2. In order for the argument to be logically sound, the meaning of "fidelity" must stay unchanged across both steps of reasoning. As long as TP and AM can be viewed as the most common interpretations of fidelity only in the context of each separate link of the RA (which I think is a fair statement), the argument appears in trouble.

4.3.2. The Ratchet Argument adopts inconsistent meanings of fidelity. What to do about it?

There are various ways to avoid or prevent this formal fault in the RA, and all have been variously (sometimes unsuspectingly) adopted in the literature. To begin with, of course, one may simply fail to notice or may deny that AM and TP are incongruous: one may think that there is only a nominal difference between the two types of fidelity (which effectively happens whenever fidelity is interpreted as a uniform notion); or that AM inherently and reliably results in TP, eschewing the burden of demonstrating a positive correlation between the two. I review examples of both strategies in next section; I think that neither (assuming an identity relation, or an inherent positive correlation between AM and TP) is sound: the challenge of bridging levels of fidelity is real. Social learning mechanisms are cognitive properties *of individuals*: they can afford descriptions only in terms of low-level, locally and temporally restricted fidelity notions like AM; any association with higher level fidelity phenomena needs explanation. Another way to avoid formal breakdown in the RA is to plug the *same* account of fidelity (AM *or* TP) in both RA claims—thus avoiding the equivocation that renders the argument unsound. Briefly and for completion, holding TP fixed across the argument leads to the following situation:

First option:

Keeping the high-level account of fidelity qua TP fixed across the RA links:

- L1. Imitation is conducive to or uniquely apt for trait preservation.
- L2. Trait preservation is necessary for cumulative culture.

L3 clearly follows and the argument is now technically sound. Suppose, furthermore, that one picks for this outline the most general meaning of "imitation", i.e. imitation as indiscriminate "all copying" (as per chapter 2): the argument looks now plausible, but loses justification and much of its explanatory bite (being now virtually tautological). The point of the RA is to single out imitation (or other specific mechanisms like language, teaching-mediated learning or "shared intentionality" abilities) with respect to other social learning mechanisms, thereby explaining a very peculiar type of massively cumulative cultural phenomenon. But this line of reasoning achieves neither the first nor the second objective: imitation as broad, indiscriminate copying is just a synonym of "social transmission and trait preservation", which may incorporate a variety of learning mechanisms; and the phenomenon purportedly explained by trait preservation, quantitative cumulative culture, is conceptually little different from an "inflated" notion of plain culture, and from social transmission itself. In other words, as said of Jablonka account, quantitative cumulative culture is just *more* simple cultural traits adding up in a repertoire (i.e. there is a quantitative, but not a qualitative, conceptual difference); but given that we established that simple culture is no different from trait transmission, and trait transmission is no different from trait preservation, the argument appears valid but minimally informative (in fact, this interpretation of the argument does not

resolve, but rather *dissolves* the explanatory gaps of L1 and L2 in a periphrastic equivalence of terms). In effect, this is perhaps the least interesting reading of the RA, which I will christen the "quantitative interpretation" (see Ch. 5). Lewis and Laland's work nevertheless doesn't fall far from it: it certainly validates L2 in a broad sense (which does not fully cover the qualitative account of cumulative culture); but it remains completely silent on L1 and on imitation's specificity. Despite the authors' claims of "providing a formal justification for the RA", L1 is in reality beyond the reach of their tools: the very mathematical formulation of fidelity adopted in this work, by referring to the probability of acquiring discrete traits in their totality, is impermeable to any integration with the social learning literature (which discriminates mechanisms mostly on the basis of input information about *partial* aspects of a behavioural trait). As it stands, it does not and cannot model the fidelity contribution of any *specific* mechanism of social learning, but must be assumed to represent just the general phenomenon of conservative social transmission.

If we reject the generalist interpretation of imitation as "all copying", and instead select any other meaning of it as "specific copying", we are thrown back to the problem of bridging levels (and filling the explanatory gap in L1): how are the low-level features of the specific mechanism of imitation (however defined: privileging the morphological, causal, or intentional source of information; being semi-automatic or under high-level cognitive control) supposed to contribute to trait preservation?

We may be luckier with the second permutation:

Second option:

Keeping the low-level account of fidelity qua AM fixed across the RA links:

- L1. Imitation is conducive to or uniquely apt for producing action-matching.
- L2. Action-matching is necessary for cumulative culture.

Again, L3 can be made to follow and the argument is sound, but we are saddled with the problem of explaining how exactly AM is supposed to contribute to cumulative culture in L2—again suspiciously similar to the problem of bridging levels. TP must be accounted for in some way in all cases, for we established that TP is a non-negotiable, basic condition for culture (it is cultural transmission in a nutshell). If we assume that the main explanatory burden for TP falls again mostly on AM, we backslide in the bridging levels dilemma, as just said: the lowlevel features of a specific social learning mechanism appearing in L1 (AM) must be brought in to explain the high level notion of TP. In the next section, I will show how this "bridging levels dilemma" has often been denied and avoided (leading to poor results), but also successfully addressed. The solution will come mostly from expanding the explanatory apparatus for high-level fidelity: it will turn out that TP and trait distributions in a population do not depend merely on low-level fidelity potentials of social learning mechanisms, but on other features of these mechanisms, as well as, crucially, on factors external to social learning, such as, for example, demographic factors, which intervene to stabilise cultural transmission.

However, there is another option connected to making AM the central account of fidelity, which delivers a very common interpretation of the RA— albeit one that is rarely recognised and motivated explicitly. I will call this the "gradualist interpretation". Under this interpretation, AM is assigned (often implicitly) a rather different explanatory role, which eventually reconnects AM to the qualitative interpretation of cumulative culture, by making it relevant to the correct acquisition of technological traits in particular. As we will see, AM acts as the cultural avatar of a specific inheritance model, which uniquely allows the evolution of cumulative, qualitatively complex and functional artefacts and procedures in virtue of it extremely high fidelity. This interpretation has its own faults and is admittedly restricted in scope, but is perhaps closest to Tomasello's original suggestion and may explain the extraordinary appeal of the RA over time.

4.4. Reconciling L1 and L2: bridging the gaps between levels

It is not entirely fair to hold that such discrepancy between levels of fidelity in RA reasoning has gone entirely unnoticed: a number of important debates have targeted the question, and the idea reliably reappears in the specialised theoretical literature on the RA and in foundational discourse on cultural evolutionary theory, as I shall examine later in this section (Henrich & Boyd 2002; Henrich, Boyd, & Richerson 2008; Claidière & Sperber 2009). However, the idea seems to have failed to make its way consistently into the less specialised literature. Works reporting the RA orthodoxy appear to routinely neglect or overlook the significance of the issue,⁴⁵ considering the difference between AM and TP a mere terminological quibble: after all, fidelity represents similarity measures in both cases, only applied at different levels. So the "scaling up" of similarity measures between the local/synchronic level of AM and the global/diachronic level of TP is (when noticed at all) often interpreted as a mere difference in terminology. But this lack of recognition contributes to the perpetuation of an oversimplified and fundamentally flawed depiction of the RA, and is particularly worrisome when the argument is taken as the conceptual foundation of empirical investigations.

In the following sections, I do three things: I first examine examples of studies that have systematically neglected the discrepancy of fidelity levels. I then present a few arguments that conclusively show that fidelity qua AM and

⁴⁵ It may also be mentioned that Tomasello himself suggested such linear connection between lowlevel fidelity of social learning and high-level fidelity of cultural transmission in one on his early papers. The portrayal of chimpanzees as preferential emulators, and the connected hypothesis of their limited access to cultural traditions because of the inherent low fidelity of emulation, has had an important influence on the field and on the idea of seamless translation or equivalence between lowlevel and high-level fidelity (Tomasello 1999).

fidelity qua TP cannot be considered equivalent or easily correlated. Finally, I present a positive case of explicit recognition and attempt to bridge the gap between low and high level fidelity, which, however, doesn't lead to a simple resolution of the RA line of reasoning, but rather to a duplication into two different interpretations of the argument.

4.4.1. Recent work on the Ratchet Argument neglects the levels discrepancy

Examples of such neglect and ambiguity about the levels of fidelity can be found easily in the applied literature, but the case of Nielsen and Tomaselli (Nielsen and Tomaselli 2010), who investigate over-imitation in Kalahari Bushmen children, will serve as paradigm. In this work, over-imitation is considered a type of extremely pedantic morphological imitation, involving the reproduction of morphological details of behaviour down to causally irrelevant or unnecessary actions.⁴⁶ The work clearly fits in the RA framework: it purports to provide empirical support for the hypothesis that striking differences of human and nonhuman cultural repertoires are due to possession of a uniquely human social learning mechanism: over-imitation (Nielsen & Tomaselli, 2010: 735). The overimitation test in this work is unambiguously based on the AM criterion, which is repeatedly described as the "fidelity" of over-imitation, and tested by comparing action similarity in the model-observer pair.

⁴⁶ For example, they introduce in the demonstration of the experimental task (opening a box) one completely irrelevant action (e.g. tapping the top of the box or sliding it in a circle) and one causally relevant but unnecessary action (e.g. opening the box with a stick, when opening with hands was faster and more efficient). They then passed on the stick and box to the child and said "your turn", with no further verbal instructions.

Children in the demonstration condition did not exhibit the target *actions* ... the children's behaviour is consistent with studies documenting children's strong motivation to *replicate with high fidelity* the object-directed behaviours of others. The *similarity* of the children's responses ... attests to the pervasiveness of this behaviour. (My italics, ibid: 732)

At the same time, however, "fidelity" is described as a feature of intergenerational transmission of skill (i.e. as TP), directly promoted by or even coincident with the "high fidelity" of over-imitation:

Other animals use tools and may have the rudiments of culture, but no animal uses tools or has developed culture with the breadth and complexity of the human species. Overimitation is a mechanism for the rapid, *high-fidelity intergenerational transmission of tool-use skills* and for the perpetuation and generation of cultural forms. The study of this behaviour promises to provide critical insight into the development of these two core human traits. (My italics, ibid: 735)

Notice also the prominence of human uniqueness claims, in typical RAorthodoxy reporting style.

The same conflation of levels of fidelity reappears in a recent body of microsocieties studies, aiming explicitly at testing the RA connection between imitation and cumulative culture in lab experiments (Wasielewski 2014; Caldwell and Millen 2008a,b; 2009). In these experiments, participants are divided into groups and instructed to complete manufacturing tasks (such as building paper airplanes, spaghetti towers, or weight-bearing devices from clay and reed) via different means, involving alternatively verbal instruction, exposure to finished products, or observation of a model building the object—in short, ideally, via different social learning mechanisms. Collectively, such studies have provided contrasting results about the importance of imitation, defined in the classic
sense of copying action morphology, actions and outcomes rather that outcomes only, or process rather than product (in short, according to the popular imitation vs. emulation distinction, popularised by Call & Carpenter 2002). What matters here, however, is that these experimental studies appear to assume a direct relation between local mechanism fidelity and global transmission fidelity. This is particularly evident in Wasielewski (2014):

Imitation, which is defined as behavioural replication in an observer following witnessing of that behaviour in a demonstrator, is proposed to be especially important for the accumulation of cultural information because it allows high-fidelity information transfer between individuals. This hypothesis suggests that in the absence of a high-fidelity transmission mechanism, information is likely to be lost; individual modifications will not be socially acquired and therefore not be subject to the additional modification that enables cultural complexity. (Wasielewski 2014:162)

In various places it emerges that the author considers "degree of bodily actions matching" (ibid.: 164) as the fidelity associated to imitation, which is supposed to have a seamless bearing on "the high-fidelity transmission [that] is important for CCE [i.e. cumulative cultural evolution]", or even more starkly to be "a necessary prerequisite for cumulative cultural evolution according to the imitation hypothesis for CCE described above" (ibid.: 162). Most of the formal and theoretical research on the topic cited in this work is equally ambiguous on the topic of fidelity measures in culture.

Though noticing this ambiguity in recent experimental work on cumulative culture, I do not mean to undermine the interest and significance of such an approach: but it should be clear that whatever information we obtain about the "fidelity" of social learning mechanisms at a local level (AM) does not automatically translate into the "fidelity" required for higher-level cultural

phenomena (TP), for there is *no* smooth, linear connection between levels, and no possibility of seamlessly translating measures of fidelity across them.

4.4.2. Action matching is not trait preservation and doesn't naturally correlate with it

One may think that this reading of the literature is uncharitable. A connection between low-level and high-level fidelity *does* exist: after all, it seems indisputable that by copying every one of Ina's movements in manufacturing a hammer stone at a very fine-grained level (AM), Abu makes his acquisition of the hammer stone-making trait more probable. While *some* connection between the reproduction of similar movements at a low level and the reproduction of a trait at a the global level obviously exists, and trait transmission could not fundamentally happen without some low-level conservativeness of the social learning mechanism, remember that we are testing here a much more specific hypothesis. The RA holds that high fidelity of imitation (AM from L1) is *necessary* for, or *particularly* conducive to, trait transmission in L2.

There are various reasons to think that this is not true: various empirical arguments show that trans-level equivalence or direct conversion of AM into TP is a faulty assumption.

First, it seems that in many cases TP can be obtained *without* AM; imitation would seemingly become superfluous. Trait transmission obtained through emulation is perhaps the argument most commonly cited by "imitation sceptics", and has represented a powerful line of objection to RA reasoning (Jablonka 2002). More generally, this seems due to a phenomenon called "motor equivalence" or "degree of freedom problem", which in brief states that there are multiple ways for an organism to perform a movement to achieve the same outcome, and no univocal correspondence exists between the motor problem posed by a task and its solution (Bernstein 1967). Suppose that I want to learn

how to drive a car from you (one paradigmatic trait hypothesised by Enquist and Laland's model); it seems unnecessary and even unproductive, in many instances, to copy every single movement you do—as fidelity qua action matching would require. I might be more comfortable activating the indicators with my thumb rather than index finger; I might prefer leaning towards the steering wheel rather than slouching back in the seat. Despite these differences, I might be perfectly able to inherit the car-driving trait from you. Likewise, it is often noted that most if not all animal traditions emerge through TP and resulting cultural transmission, in the complete absence of the AM associated with imitation (Brown 2017). A number of other examples may be formed in support of this thesis: so AM often seems unnecessary to reproduce a variety of behavioural traits.⁴⁷

Second, AM might be neither necessary nor sufficient for TP. One may think that despite the motor equivalence argument, matching precisely a model's actions is certainly a safe (if not the only) strategy for learning the trait (i.e. sufficient, if not necessary). This, however, doesn't seem to be the case: putting too much effort into copying the exact movements of your model might subtract attention from the achievement of the important sub-goal necessary to the successful acquisition of the trait, from understanding complex causal passages of the procedure, or from memorizing the general structure of behaviour—all factors that may be more crucial to the overall acquisition of the trait. More generally, the argument applies whenever cognitive resources necessary for the faithful reproduction of action morphology (AM) trade off against the maintenance of the trait in its entirety (TP).

A similar and important counterargument to the relevance of AM to TP appeals to the difference in physical resources between the model and the observer. If we hypothesise that, especially in traditional societies, the learning cohort will often be comprised of younger subjects, who are typically less strong,

⁴⁷ However, on the other hand, there are important exceptions to this objection: these will be the fuel for the gradualist interpretation of the RA.

fast and coordinated than skilled and fully developed adults, attempts to copy the exact actions of the model might (AM) be counterproductive in the acquisition of the relevant behaviour (TP). This is true in general when the physical capabilities of the observer are significantly different from those of the model. (It may even be that the physical equipment of the learner is superior to the model's one for the task at hands, making AM superfluous or again purposelessly inefficient for the learner: whenever a task is better accomplished by individuals of smaller size—probing a burrow, for example). Significant differences in height, size, strength, limb length, can make particular movements suboptimal for a subject and optimal for another.

Suppose, for example, that I am learning how to build a bow from you, and that I am significantly shorter. Your technique for bending the wood length is to place one end of it in between your feet, holding it firmly, and to pull down the opposite end to fasten the string. Faithfully copying your technique (in the AM sense) might be unsuccessful in my case, because I might not be able to reach the other end of the shaft; I might be better off to place one end in a corner or against a fixed support and bend the length by moving closer to the anchoring point.

This last argument may also have a substantial impact on the motivation of the learner: being unable to achieve the relevant results through pedantically copying the actions or process, learners might soon lose interest in attending to the model. Again, this seems a plausible objection to raise again particularly for traditional learning settings, where youngsters typically learn from experienced models. The absence of motivation to imitate, however, cuts the possibility of trait transmission at its root: if learning by imitation is too wearisome and inefficient, and thus fails to be reinforced by reaching the desirable outcome of behaviour, imitation again incurs the flop problem.

All of these arguments show, at the very least, that AM doesn't translate automatically into TP, and that therefore the two meanings of "fidelity" associated with the two phenomena should not be considered equivalent, nor

205

easily correlated.⁴⁸ The problem of bridging levels is genuine and needs to be addressed. I now discuss the work of two authors that have taken up this challenge, and the significance of their results.

4.4.3. One proposal to bridge the gap between low-level and highlevel fidelity; a third account of fidelity emerges

The most prominent and explicit attempt to directly tackle the problem of fidelity at different levels is due to Henrich and Boyd (Henrich & Boyd 2002). In their 2002 paper, these authors endeavour to present a direct response to Sperber and other cognitive anthropologists' attack to "neo-Darwinian" approaches to cultural evolution and to their presumably misguided assumption about high fidelity cultural transmission.

As we saw before, the cognitive anthropologists' crucial accusation is that cultural transmission is habitually low fidelity and imprecise, and therefore that cultural evolution cannot not be modelled after high fidelity replicator's dynamics derived from the biological examples of genes. Perhaps contrary to expectations, Henrich and Boyd (who according to cognitive anthropologists are among the sinful ranks of the neo-Darwinians) do not reject the accusation: instead, they embrace the idea that cultural transmission (at least in certain cases, ibid.: 89) is on average very imprecise and noisy *at a low, individual level*; but not so *at the higher, populational* level. Here, they claim, replicators' models still represent the right heuristic: cognitive anthropology's mistake is "to assume that the only process that can give rise to accurate replication at the level of the population is accurate replication at the level of individuals" (Henrich & Boyd 2002: 99-100).

⁴⁸ Some of these arguments against the equivalence of AM and TP can be found in Sterelny 2012; others emerged in personal conversations with the author.

The emergence of high fidelity at the populational level is due, they argue, to additional factors that intervene to stabilise representations and reduce noise from the individual level. Among such factors, Henrich and Boyd enlist cognitive attractors, thus essentially ratifying the picture proposed by cognitive anthropologists—at least at the low level. The fact that some representations are "easier to think" makes individuals converge towards them (ibid.: 92). In addition, however, they suggest that conformist and prestige biases in transmission (that is, the empirically documented tendencies to preferentially copy the most frequent behavioural variants, or prestigious individuals) play a crucial role in gathering scattered individual-level representations into uniform and discrete population-level representations or traits. Demographic factors such as the size of the cultural learner's pool are also important to compensate for errors in transmission and stabilise high-level flows of cultural information.

Introducing these biases in formal models where individual fidelity is low and traits fail to be transmitted in almost half of the cases, they show how these propensities "correct" the tendency of the system to "rapidly converge to a random distribution of mental representations" and lead instead to cultural inertia and cumulative cultural adaptations (ibid.: 90; 97; 99). Here a little terminological clarification is in order: "cultural inertia" is just another way (modelled after the analogous expression "genetic inertia") of saying trait preservation, or trait longevity, or trait transmission, or simple culture: it captures the idea that cultural traits, by being learned and re-transmitted, simply stick around in a recognizable form. As far as genetic inertia aka cultural transmission is concerned, we may even say that the positions of Henrich and Boyd on one hand, and Sperber and the cognitive anthropologists on the other, are highly convergent. Essentially, both parties recognise that poor fidelity at a low level can transpose into significant fidelity at some higher level, given the right conditions and stabilizing factors.

Notice also that, after these clarifications, it becomes clear that the cognitive anthropologists were *never* against fidelity *in toto*: by recognizing that

207

cultural transmission happens (which they undoubtedly do), they must accept that some fidelity qua TP is in place. The very fact that we can say that the Little Red Riding Hood tale *is* transmitted rather than not means that the tale's representation that you acquire from me *resembles* mine in some crucial respect: that is, there is some important degree of *similarity*, or fidelity. So their fidelity scepticism could concern only low level, AM-like fidelity.

As can be seen, terminological questions are insidious in this area, and disambiguation is important: the fact that cultural anthropologists prefer the label "cultural transmission" to describe what others have also alternatively described as "TP-fidelity" has fooled us (and arguably themselves) into thinking that Sperberians are *full* fidelity sceptics. But they are not; and they cannot be, at least if we accept Laland's picture of fidelity qua TP.

Where disagreement re-emerges between the two parties is, however, in the explanation of high level cultural dynamics: for Henrich and Boyd, the fact that high level transmission can be redescribed as high-fidelity transmission justifies the application of a selectionist framework at the high level. Discrete high-level cultural representations, they claim, can now go back to compete against each other like good Darwinian individuals and get selectively retained. Cognitive anthropologists and attraction theorists disagree: they maintain that selection can at best explain a marginal fraction of existing cultural representations, their frequency and distribution; the bulk of these phenomena is still explained by cognitive attraction, i.e. by our inescapable gravitation towards cultural traits that better fit our minds.

I do not wish to take a side on the selection/attraction debate—and in any case settling this question largely exceeds the scope of this chapter and ultimately of the entire work. Briefly and for completion: I think that neither framework is *generally* superior; rather, each can have its explanatory returns and advantages, depending largely on the cultural phenomenon we wish to explain. Recently, it has been suggested that the main difference between cultural attraction theory and mainstream cultural evolution is one of explanatory targets: while the former focuses on fully human, modern, often abstract, linguistically- and symbolically-mediated cultural phenomena, the latter privileges the explanation of functional, fitness-laden, generally more ancient and to some extent phylogenetically generalizable cultural adaptations (Sterelny 2017). I think this suggestion really gets at the root of the disagreement and dissolves it as mostly apparent. Perhaps a fruitful direction for future research would be to spell out more precisely and analytically *what* features of the cultural phenomenon in question makes one explanatory framework more suitable than the other, rather than fuelling the squabble about who gets it fully right (a direction I have very modestly attempted to take with the discussion of efficiencies in chapter 3).

Moving on from the selection/attraction debate, and returning to the overarching task of disambiguating the RA, Heinrich and Boyd's account may ultimately be held to deliver three important messages.

The first and self-evident message is the ratification and problematisation of a level-sensitive picture of fidelity. Surprising as it may sound now, the literature is lacking in explicit acknowledgments of such level-sensitive nature; even less frequent is the recognition that levels of fidelity are genuinely irreducible to each other, and low levels cannot be assumed to flow unproblematically into higher ones. Heinrich and Boyd are very valuably forthright in this respect.

The second message is that, to the extent that Boyd and Heinrich's response to attraction theorists can be read as RA-relevant (and I think it was definitely advanced in the spirit of defending a central RA tenet), then the RA really is a protean entity, lending itself to a range of interpretations. The one these authors have in mind, for example, clearly diverges from others in various respects. From the elucidation of Boyd and Heinrich's interpretation, finally, I draw a third and crucial message about the quality and quantity of fidelity levels at play in the RA conceptual space.

Let us start from the second message: I will briefly tie Henrich and Boyd's considerations back to the RA and look at their significance for the argument. The first thing to notice is the clear appreciation that cumulative culture explanations support two levels of fidelity, which the authors identify as individual and populational. Are these levels explanatorily relevant to RA-like constructs, though? I think they are, although they eventually come to portray an abridged RA at best. The significance of imitation to the argument is in fact largely downplayed: the authors accept that in many cases there is no low-level fidelity—at the level of local social learning. In this sense, to the extent that we wish to consider the RA as integrally made up by the combination of L1 and L2, Henrich and Boyd's reasoning departs substantially from it: by dropping imitation, they essentially eliminate L1 from the argument. Surely we now have an interpretation that connects the *lack* of fidelity at some low-level to the presence of fidelity at some higher level (see below), which represents a genuine advancement and a clear appreciation of the levels ambiguity in fidelity. But in its entirety, these conclusions also embody a considerable steering away from the full set of RA inferences.

Instead, we may see Henrich and Boyd's reasoning as a robust endorsement of L2 of the RA: the centrality of fidelity is first salvaged and then strongly reaffirmed. With respect to other classic interpretations of the RA, however, the essential role of fidelity is escalated to the collective level. Let us look in slightly more detail at what this kind of fidelity is thought to explain, and how.

As mentioned earlier, the authors maintain that re-establishing faithful transmission at the high level opens the way to explaining two important phenomena: *cultural inertia* and *cumulative cultural adaptations*. What is the relation of these two phenomena to cumulative culture, the original RA explanandum? Roughly, I think that cultural inertia can be assimilated to simple culture or *quantitative* cumulative culture (Jablonka's notion), while cumulative adaptations retrace more closely the notion of *qualitative* cumulative culture

210

(Tomasello's notion). Furthermore, it seems reasonable to presume that a difference exists between cultural inertia and cumulative cultural adaptations, or else the authors would not bother indicating them as two distinct explanatory targets. This explanatory connection of high fidelity and cultural inertia is, I think, another case of dissolving rather than resolving L2 under one possible interpretation of its terms—it is very similar again to Laland's or Jablonka's quantitative interpretation of the RA. "Cultural inertia" is essentially another term for "cultural transmission", simple culture and quantitative cumulative culture à la Jablonka: whether one chooses to conceptualise it at the level of individual traits, or at the level of trait distributions, cultural collectives and population repertoires make little difference. If we buy into this equivalence (and I do not see exactly how the various terms can be sharply conceptually distinguished), again, I would say that even fidelity at the collective level does not so much *explain* these phenomena as *coincide* with them or merely describe them.

Perhaps a more informative explanatory path is proposed to connect high fidelity to cumulative cultural adaptations. As I noted, we might feel entitled to presume that, roughly, by "cumulative cultural adaptations" Henrich and Boyd intend something akin to *qualitative cumulative culture*. Adaptations, in fact, are typically represented as traits that increase the efficiency and fitness of organisms in response to environmental challenges; this notion appears to have clear affinities to the increase in efficiency that was recognised as a distinctive trait of qualitative cumulative culture. Adaptations, however, are also by definition a product of natural selection: what Henrich and Boyd are suggesting, by postulating cumulative cultural adaptations as the explanandum, is that fidelity must have a crucial role in explaining their emergence *in a selective framework*. In effect, RA ideas are not extraneous to selectionist considerations; however, in most current interpretations these have slid back into a secondary position and more often than not they are not explicitly referenced. Henrich and Boyd's reading, instead, places a bolder selectionist slant on the argument. Note

211

that, by looking at fidelity of repertoires, Heinrich and Boyd's reasoning appears to presuppose a special brand of selection mechanism at play: cultural *group* selection, acting at the level of collective adaptations (see StereIny 2017 for a discussion of plausibility).

However, if one were to search Henrich and Boyd's paper for a more explicit account of *why* exactly selection needs high fidelity transmission to produce cumulative cultural adaptations, one would be disappointed: the authors just refer back to the biological argument about high fidelity genetic transmission, which alone "can generate cumulative adaptation" (ibid.: 98). So while L2 of the RA receives resounding support from these authors, their commitment to providing a well-detailed content to it still wavers. They offer a hint, though: the standard reference to biological arguments about fidelity and cumulative selection. In the next section, I finally look at the biological arguments connecting high fidelity to cumulative adaptations, which seem to constitute the fundamental analogy on which L2 of the RA is built.

Before I move on to this task, however, I shall point out a final feature of Henrich and Boyd's analysis, which is particularly relevant to the purpose of this chapter. Regardless of their exact role in a RA construct, the notions of fidelity proposed by these authors seem at odds with those previously isolated in the work of Sperber and Laland. As I have remarked, Henrich and Boyd's work shows a clear appreciation of two levels of fidelity: individual and populational. These two levels can certainly be characterised as high and low, but notice that they don't match the high/low distinction that I identified earlier for Sperber and Laland's fidelities. The lower level associated with AM and Sperber's account is somewhat *below* the individual level: it is a *sub*-individual trait level, which appears to discriminate fine-grained components, details or parts of single traits. It has much higher resolution, so to speak, than the individual level. Likewise, the higher level associated with TP in Laland's interpretation comprises the individual and the populational levels (which instead are, respectively, Henrich and Boyd's low and high level). So there are at least *three* levels of fidelity in RA discussions, not two: the "low" level can in fact be interpreted as the individual trait level (as Henrich and Boyd do) or as a *sub-individual* level of trait components, like the one considered by fidelity qua AM. From now on, I will summarise these three levels as follows:

- Fidelity-I: transmission, similarity, inheritance, preservation of sub-components of individual traits; AM is an instance, designed specifically for topographical details of behaviour.
- Fidelity-II: transmission, similarity, inheritance, preservation of individual traits; exemplified by TP.
- Fidelity-III: transmission, similarity, inheritance, preservation of traits distributions in populations. Exemplified by the preservation of repertoires and traits distribution in populations.

As we have seen, the usage and conceptualisation of fidelity in the RA is quite liberal; even when level transitions are problematised, like in Boyd and Heinrich's work, the picture may remain partial, tracking some levels but omitting others. A systematic and well-organised conceptual space for the notion seems largely missing. The tripartite distinction obtained here is far from offering a robust and detailed solution to this problem; however, it may represent a sensible starting point for future conceptual clarifications.

What the tripartite distinction can certainly do for us, instead, is to clarify the inconsistent usage of the term in RA contexts. The notion of fidelity is not the only aspect of the argument that is largely perceived as univocal; even when it is not, and when level-savvy analysis is introduced, the *transition between levels* is generally represented as smooth and unproblematic—just a matter of enlarging or reducing the grain of analysis. I do not think, however, that this is correct. The message from Henrich and Boyd's work, as well as the one from the non-obvious correlations between AM and TP (4.4.2) is precisely this: there is no obvious and straightforward connection between levels. Just as in some cases low levels of

fidelity-II (TP) do not correspond to low levels of fidelity-III, in other cases high levels of fidelity-I (AM) do not correspond to high levels of fidelity-II (TP). We just cannot assume that significant (or poor) fidelity at lower levels will translate into the equivalent amount of fidelity at higher levels. (The reverse is also probably true, but generally less relevant to RA considerations.) There may be cases of positive correlations, undoubtedly; but these must be actively documented and explained, rather than assumed.

So, what for the RA? Heinrich and Boyd's considerations have uncovered a feasible bridge between low fidelity-II and high fidelity-III, by appealing to a host of scaffolding mechanisms like learning biases, attraction-like forces, and demographic factors. But this solves a quite different problem than bridging the gap between fidelity-I and fidelity-II, which seems to be a central concern of many other RA readings. In light of Boyd and Heinrich's results about the relative unimportance of TP, in addition, the insistence on such readings of fidelity qua AM becomes almost puzzling.

The next three sections look at answers to this puzzle. In order to motivate the insistence on the finest-grained brand of fidelity, fidelity-I or AM, we will need to go back to the biological root of arguments about fidelity.

4.5. The argument from biology: fidelity as inheritance

Earlier I mentioned that the fundamental meaning of "fidelity", level controversy apart, is not particularly contentious: fidelity should be interpreted as a measure of similarity. One important reason to think so comes from examining the notion of fidelity in the biological context. The biological context represents the archetypal analogue for evolutionary reasoning in the cultural domain, via the connecting route of so-called substrate-neutral accounts of evolution (i.e. the idea that evolution is essentially an algorithm that can be applied to different substrates, biological and otherwise—also expressed as "universal Darwinism" (Dawkins 1983; Dennett 1995). Thus, although there are good reasons to be wary of analogical reasoning and of "direct imports" from the native to the derived field, biology is without doubt a good place to start to understand the significance of fidelity in cultural evolution (Godfrey-Smith 2009).

In biology, fidelity is routinely and most naturally discussed in relation to DNA replication (Loeb & Kunkel 1982, Kunkel & Bebenek 2000, Kunkel 2004, Fernandez-Leiro, Conrad et al. 2017). DNA replication occurs in all living organisms before cell division, leading to the formation of two identical (or nearly identical) DNA molecules from a single parent-molecule, and it is commonly presented as the basis of biological inheritance—by allowing each cell to host a replica of the whole organism's genome. The attribution of high fidelity to DNA replication suggests both its fundamental meaning and its central context of significance—that is, the biological phenomenon of inheritance.

Let us have a closer look at why DNA replication is considered the quintessential example of high fidelity by quickly reviewing the fundamental steps of replication processes. When replication begins, DNA strands are unfolded from their stable double-helix configuration, so that each strand can serve as a template for the production of a new counterpart. The counterpart strand is built by attaching new nucleobases (A, C, G, T) to the template strand. The addition of nucleobases follows a rule known as "complementary base pairing": successful pairings can occur only between A and T, and C and G. Base-pairing is said to be complementary because the binding sites of, respectively, A and T, and C and G, "complete" each other to form stable hydrogen bonds, with two bonds formed between A and T, and three between C and G: this process is generally very accurate, meaning that aberrant pairings (say A-G or T-C) occur with extremely low frequency. The standard complementary base-pairing rule, however, can fail on some occasions. For example, with slight changes in helix geometry, two hydrogen bonds can form between G and T. Likewise, but without any change in helix geometry, rare tautomeric forms of the four DNA bases (which occur fleetingly in the proportion of 1 every 10⁴ or 10⁵ bases) can mis-pair: for example, the rare tautomeric form of C pairs with A instead of G.

The non-standard bonds defying the A-T/G-C rule are chemically less stable than the standard ones: they are "wrong" pairings. However, a host of proofreading and repairing mechanisms (e.g. DNA polymerase), which are being increasingly studied and documented, intervene to sequentially correct such mistakes and replace the "wrong" nucleotide with the "correct" one (Alberts, Johnson et al. 2002, Pavlov, Frahm et al. 2006; Albertson & Preston 2006).

Each of the two new DNA molecules resulting from this process ends up containing a strand from the old molecule and a newly formed strand—virtually identical to the one formerly attached to the template strand if the new bases are in the "right" loci. The process is known as *semiconservative replication* (Meselson & Stahl 1958), and occurs both rapidly (at the rate of several hundred new nucleotides per second) and with astonishing *fidelity*: only 1 mistake is made for every 10⁹ nucleotides copied, so that on average less than one mutation per genome per cell division occurs (Drake, Charlesworth et al. 1998).

What is meant by "fidelity" here is reasonably clear: fidelity means *errorless*, or error minimizing, replication. The interesting question, hence, becomes what counts as error, and how is it calculated? I think we have two possible rough answers to the question of what counts as error in replication, but only one of them counts towards determining fidelity. On one hand, flawed, non-standard base pairing occurrences could be considered as errors, being typically less stable and chemically deviant. Here "error" means malfunctioning of the mechanism. On the other hand, one could compare the final DNA molecules and spot mismatches in the newly created strands—i.e. a different nucleotide sitting at a specific locus in the new chain, e.g. A where it "should" be G-by comparison with the nucleotide sitting at the corresponding locus in the old chain. In this case, "error" means mismatch. Deviation from an expected norm defines error in both cases; but "norm" can be the standard chemical behaviour of the involved molecules or the expected result defined by the mother DNA strand. The second type of "norm" is defined by the pre-existent mother strand: it is fulfilled when the newly created DNA strand is identical (or as similar as possible) to the strand that was previously coupled to its template strand.

One may object that the only reason that we judge a chemically aberrant G-T pairing as "wrong" is that it is *dissimilar* from the previously occurring or ideal (in the sense of archetypical, platonic) pairings A-T and C-G that embody the pairing rule, and thus that there is no actual difference between errors as mismatches and errors as malfunctioning. But this is not correct. The reason that we consider mispairings "wrong" has more to do with the fact they realise less-than-optimal, unstable chemical bonds than with the fact that they don't resemble previously occurring or "ideal" ones; conversely, the reason we judge mismatches "wrong" has more to do with the failed similarity to the mother strand than to the underlying malfunctioning of the mechanism.

The question is now *which* kind of error counts towards fidelity measures. In DNA replication these two types of error co-occur, that is, malfunctioning produces mismatching—unless correction mechanisms intervene; but *in principle* the two types of error could come apart. I think that the relevant type of error for fidelity measurements is mismatching, *not* malfunctioning.

To see why error-as-malfunctioning is fundamentally unlike error-asmismatching in DNA replication, and why only the second type counts towards fidelity measurements, consider the following quick thought experiment. Suppose an imaginary experimenter E is tasked with establishing the fidelity of a specific DNA replication event, which she can closely observe as it unfolds, i.e. tracking every single nucleobase addition to the template strand. Suppose also that E works under the assumption that fidelity means errorless replication, but under no specific assumption about the meaning of "error". E observes replication and counts, say, three errors—that is, three malfunctionings of the complementary base-pairing mechanisms that also fail to be corrected by DNA polymerase proofreading, and thus result in three mismatchings in the final strand. E rates the replication process as poorly accurate and duly annotates a report with her observations, which is passed on to a second experimenter E₂, who is in charge of reassessing the result and of synthetizing more copies of the molecule. Unbeknownst to both experimenters, while the molecule is being handed down together with the report, one of the traditional impertinent demons of thought experiments intervenes to restore order and replaces the "wrong" nucleotides with the "right" ones. When experimenter E₂ assesses the target molecule, he finds that it is identical to its mother molecule—at the same time, he can read on the report about the malfunctioning recorded by E. Would E₂ rate the new molecule as a high fidelity copy of its predecessor or not? I think she would. Model-copy correspondence would override malfunctioning in the judgment of errors relevant to fidelity. If a deviation from the expected functioning of the replication mechanism occurs, we are also naturally led to speak of some kind of "error" or inaccuracy. But this kind of error has little to do with the calculation of fidelity.

Time to sum up the primary message from biology. "High fidelity" in biology is most commonly encountered in the description of *error-free replication* of DNA, where "error" means mismatching. Replication is a process that is expected to produce a *copy* from a pre-existent original (a molecule, in this case). Whether the copy is a good or a bad one is defined by its *similarity* (i.e. matching) to the parent molecule. Thus, ultimately, "fidelity" is defined as a measure of *similarity* degrees. It is high when similarity is substantial or complete (and error low); it is low when similarity is poor (and many dissimilarities are registered); it is null when no similarity at all can be detected. In turn, similarity in evolutionary processes chiefly concerns the phenomenon of inheritance.

4.6. Models of inheritance: replication vs. reproduction plus inheritance

The received view on biological inheritance is genocentric, and states that organisms resemble their parents in virtue of the *replicated* genetic material the latter hand down to them. However, such a genocentric view has been increasingly challenged by and progressively integrated into what one may call "the epigenetic or multi-systemic view" of inheritance. Multi-systemic views postulate a variety of extra-

genetic "inheritance systems" that, alongside the genetic one, can contribute to parent offspring similarity (Jablonka & Lamb 2005; Mameli 2004; Odling-smee, Kendal, & Tehrani 2011; Odling-Smee 2007; Sterelny 2001). Such a view defines inheritance as a process whereby an entity (the "parent") affects another entity (its "offspring") in such a way that the offspring *resembles* its parent more than other members of the population, with respect to a given trait (Mameli 2004; Godfrey-Smith 2009). Cultural inheritance, for example, fits into this framework identifying teachers or models as "cultural parents" and learners or observers as their "cultural offspring", even if biologically unrelated. One crucial difference in these "applied" views of inheritance concerns the degree of fidelity. While genetic inheritance is, as we've seen in the previous section, the epitome of faithfulness, in the multi-systemic perspective inheritance is a softer relation of parent-offspring similarity.

A similar situation may be observed in abstract conceptualisations of inheritance operated by different substrate-neutral accounts of evolution. Two broad, competing models for substrate neutral accounts of evolution may be individuated, which Godfrey-Smith (2009: 4) labels the "recipes" (or classic) and "replicators" accounts.⁴⁹

In the classic tradition, descriptions of natural selection tend to take the form of a recipe: given such and such "ingredients", evolution by natural selection will follow. All recipes call minimally for a relatively integrated collection of entities (a population), which must be capable of reproduction, display variation, have differential fitness (i.e. difference in reproductive output), and possess some form of

⁴⁹ There are other and more fine-grained ways of counting approaches. For example, Sterelny et al. (1996: 378) distinguish "four very general ways of characterizing evolution": the "Received View" (roughly correspondent to Godfrey-Smith's recipes); the "Gene's Eye" (i.e. Dawkins and Hull's perspective); the "Extended Replicator" view; and the "Developmental System" perspective. I see this grouping as more relevant to the units of selection debate. Others contrast the formal treatment of evolution by natural selection (e.g. Price equation) with verbal ones, such as recipes or replicators (Bourrat 2014a). However, for the purposes of my discussion, the basic distinction between "recipes" and "replicators" approaches will suffice, since fidelity plays an important role in it.

inheritance—where "inheritance" means parent-offspring similarity (ibid.: 5). Any population possessing these features will automatically undergo natural selection. On the other hand, the "replicator" approach, developed primarily by Dawkins and Hull in the 70s and 80s offers quite a different scenario and is often pictured in tension with the classic one. Emphasis is here placed on units of replication rather than on reproducing organisms, and these units are focused on their mostly competitive agendas. The replicators account is still an instance of populational thinking, but crucially, in this view, replicators are the populations *that count*: they are the necessary engine of evolution, carrying into evolutionary dynamics "the rest of the biological world" as a mere passive product of their activity (ibid.:2). The fundamental concepts of reproduction, inheritance, variation, and differential fitness are not extraneous to the replicators approach—but the fact that they are here performed by the special mechanism of *replication* makes all the difference (according to replicators enthusiasts).

One of the most obvious differences between recipe and replicator accounts concerns the amount of fidelity postulated for inheritance. In the replicator approach, which is essentially modelled on genes, inheritance coincides with replication—and as we've seen, replication here is by definition extremely faithful. By contrast, the classic account is more permissive: being modelled on organisms, it accepts that offspring are not *copies* of their parents; rather, parent-offspring *similarity* is sufficient to define the inheritance relationship.

There is another, perhaps less frequently remarked, difference between the two accounts. To see what it is, we need to take a step back and examine the kind of relationships that characterise the parent-offspring pair. The parent has a broadly causal effect on the offspring in two important and distinguishable ways. On the one hand, the parent is causally responsible for the *existence* of the offspring. In "regular" biological cases, it is sometimes said that the parent is materially/causally responsible for its progeny, for most cases of biological reproduction involve some material overlap between the two—the handing down of genetic material (Griesemer 2014, 2005; Godfrey-Smith 2009). Roughly speaking, this role coincides

with the reproduction required in classic recipes; it captures the idea that the parent entity "produces" (brings about the existence of) a new entity. On the other hand, the parent is causally responsible for the *similarity* of its offspring. This may be described as formal/causal responsibility of the parent, and, again roughly speaking, it is generally identified with inheritance. This duality in the causal role of the parent for the offspring has been remarked in numerous places under different names in the theoretical literature on natural selection (Godfrey-Smith 2009; Bourrat 2014a, 2014b; Griesemer 2005; Griesemer 2000; Charbonneau 2005).

In regular biological cases, and to some extent also in evolutionary theorisation, these two roles co-occur and are often indistinguishable: adult dogs produce relevantly similar dog puppies, not parrots or bikes. The formal responsibility of the parent in generating a similar entity regularly overwrites its material responsibility. This is also evident in the classic formulation of the concepts of *reproduction* and *inheritance*, which seem to be blends of the two causal roles, though perhaps in different proportions. "Reproduction" is regularly defined as the generation of a new entity of the *same* kind (i.e. *similar*), while "inheritance" denotes *parent-offspring* similarity (and what is parent-offspring if not the *reproductive* relationship?). ⁵⁰ The reproduction-inheritance entanglement is an extremely complex issue that cannot be satisfactorily unravelled here (see Godfrey-Smith 2009 for comprehensive discussion). Suffice it to say that the recipe approach, by distinguishing reproduction and inheritance, suggests that there are (at least in principle) *two* forms of causal influence exerted by the parent on the offspring (though they might not be easily separable in reality).

If the two causal roles of parents can be to some extent distinguished in recipe approaches, this is not the case in the replicators' framework. Replication is by

⁵⁰ In effect, *reproduction* and *inheritance* are fatally entangled concepts, leading to a rather confusing coexistence of statements. One may read that "inheritance … entails the concept of reproduction" (Bourrat 2014a: 2), alongside Darwin's famous concluding remark in *The Origin of Species* (summarizing the first two "laws" of evolution by natural selection): "Growth with Reproduction; Inheritance which is almost implied by Reproduction".

default a catch-all notion: its occurrence takes care simultaneously of "reproduction" (material/casual responsibility) and "inheritance" (formal responsibility). While it is possible to imagine a parent that produces "something else"—a different kind of offspring (albeit requiring a slightly sci-fi thought experiment)—it seems logically contradictory to imagine a case of replication that produces something different from a *copy* of the original. Such a case would not be replication. Thus in the replicators framework, *reproductive lineages* and *hereditary lineages* (so to speak) always and necessarily coincide.

Let us focus first on the different take on *fidelity* of recipes and replicators approaches. As I said, the two accounts certainly differ in their fidelity requirements for inheritance: the replicators account postulates that extremely high fidelity is necessary for evolution by natural selection, while the recipe approach is content with moderate levels of fidelity to the same end. However, unless replication has completely different properties, this difference may be rephrased as one of degree rather than kind. In other words, even though the accounts differ in the degree of similarity/fidelity required, they nonetheless agree that intergenerational similarity (regardless of its degree) is necessary.

4.6.1. The roles of inheritance and similarity in evolution

Perhaps one good place to start is to ask *why* such similarity is so crucial to evolutionary matters. The reason is profound and has to do with the way natural selection acts on populations to produce what we identify with evolution, which is, in its simplest definition, a change in the frequency of variants in a population. Let us hence revive a famous metaphor that illustrates the workings of selection. The metaphor is due to Dawkins, who likens selection to a sieve. If we scatter a bunch of (say) grains into a sieve, some will pass through and some will not, depending on the tightness of the mesh and on the size of the grains. The mesh narrowness is an analogy for selective pressures; the size of grains for the selection-prone properties

of individuals.⁵¹ The sieve separates entities possessing a target property (being above a certain size) from those that don't (being below that certain size), thus imposing a certain order to the entities. Dawkins calls this process "single-step selection" (Dawkins 1986: 45). Single-step selection is, however, a limited kind of order-imposing mechanism, and rather different from our idea of natural selection. Going back to the sieve metaphor, in single-step selection grains are sorted once and for all, and we would have to remove the original grains and proceed to a second load if we wished to sort more. Here natural selection differs from regular sieves. With natural selection, the sieve never finds itself empty. Grains that survived one round will contribute to the next load by *generating* new and *similar* grains. Natural selection is a process of iterative, self-feeding sifting, or, as Dawkins calls it, of "cumulative selection". In cumulative selection:

... the results of one sieving process are fed into a subsequent sieving, which is fed into ..., and so on. The entities are subjected to selection or sorting over many "generations" in succession. The end-product of one generation of selection is the starting point for the next generation of selection, and so on for many generations. (ibid.: 45)

So, single-step selection is static: once happened, ever happened. It is a selfexhausting rather than self-feeding process. Cumulative selection, by contrast, keeps refilling the sieve, almost miraculously (with no external intervention), making it a permanently dynamic, *iterative* system. In addition, the process is said to be *cumulative* because self-feeding re-starts from the results of the previously filtered round: whatever properties selection favoured in one round, they will be found in the next one, ready for another skimming.

⁵¹ To be more properly analogous to natural selection, our sieve should be imagined as varying the tightness and shape of the lattice over time, though at a slower pace of than that of the replicating entities; besides being sensitive to a much more varied range of multidimensional properties than size, of course.

It is fairly clear how these extraordinary features are achieved. The "grains" have been endowed with two signature properties of paradigmatic Darwinian individuals, or something akin to them: reproduction and inheritance. The sieve metaphor allows one to intuitively see the relative roles of reproduction and inheritance in the process of evolution by natural selection. Quite simply, the contribution of reproduction or material causation is that of refilling the sieve with "stuff": reproduction defies self-exhaustion and makes the process *iterative* and *self-feeding*. The role of inheritance or formal causation is subtler and more articulated.

At a very general level, it is sometimes said that inheritance is necessary to preserve the distribution of variation in the population (preventing stochastic distributions), and that variation at the populational level is in turn necessary for selection to act upon (given that selection, by definition, discriminates variants Charbonneau 2014, Lewontin 1985). More importantly, however, the role of inheritance is that of maintaining *a* particular distribution of variants in the population (rather than distribution in general): specifically, the variants that have resisted selection, by producing similar offspring, will be found more frequently in the successive generation. Thus inheritance allows the population to "track" the fitness or success of particular variants at each step (at each "sifting"). By happening iteratively, this process of "memory" or retention of advantageous variants will shift trait frequencies in the population towards local optima and can produce adaptations (Dawkins 1986; Charbonneau 2014 for the use of "memory").

This "success tracking" role of inheritance is key to the cumulative nature of evolution by natural selection. Without parent-offspring similarity and the ensuing opportunity of restarting future selection from "pre-filtered" results, things would have to restart from scratch every time, and no directionality in change would occur. Imagine, for example, that only reproduction but no inheritance were in place. In this scenario, we would perhaps still observe some sort of self-feeding and iterative process, but certainly no cumulative build-up of properties favoured by selection: their increased frequency won by resisting selection would be lost and wiped out at each round, as the process reboots.

224

(Incidentally, this last point may indicate another crucial role for inheritance, see Bourrat 2014a, 2014b. Suppose, as we did, that an imaginary population had reproduction *without* inheritance—i.e. parents generating offspring of completely different kind—and that through reproduction such a population could effectively maintain some sort of self-feeding, iterative dynamic. Inheritance is nonetheless necessary for this scenario, inasmuch as offspring would need to be similar to parents in at least one respect: the ability to reproduce.)

All these functions of inheritance that can be seen at work in Dawkins' sieve metaphor give us the basic picture of what evolution by natural selection is about. The question is whether inheritance is here defined by faithful replication or mere parent-offspring similarity, and whether this difference in fidelity degree makes any difference. Of course, the model of inheritance that Dawkins implies is replication. However, what is interesting is the sieve scenario doesn't seem to *need* replication— all the functions of inheritance enlisted above could be met, albeit perhaps at a lesser degree and with less pronounced effects, by the more modest notion of inheritance as parent-offspring similarity.

In fact, despite the enormous influence that the replicators account has had on evolutionary theorisations and on the perception that extremely faithful, genetic-like inheritance is necessary for cumulative evolution, a substantial amount of work converges towards the idea that high fidelity replication, although important, might not be *necessary* for evolution by natural selection (Godfrey-Smith 2009; Bourrat 2014a, 2014b; Charbonneau 2014; Griesemer 2005; Wimsatt & Griesemer 2007).

Taking these theoretical results back to the debate on cultural inheritance, and adding them to the perspectives offered by Henrich, Boyd and Laland described in the previous sections, the RA's emphasis on extremely high fidelity inheritance may appear even more puzzling—particularly if we consider the RA claims to concern AM, or sub-individual level fidelity. The next section finally clarifies where this requirement for extreme high fidelity comes from.

4.7. Weismannian inheritance and the gradualist interpretation of the Ratchet Argument

Numerous argumentative lines have come together in this chapter to suggest that, contrary to the RA orthodoxy, extremely high fidelity transmission or inheritance may not be a necessary condition for cumulative cultural evolution, after all. Moderate levels of fidelity, which we may associate with individual-level TP, seem sufficient for the establishment of recognizable cultural traits and stable, perhaps even complex cultural repertoires. This argumentative line may be seen as corroborated from a number of directions, including cultural attraction hypotheses, Laland's quantitative accumulation hypothesis, as well as from all the "imitation sceptics" and defenders of animal cultures, such as Jablonka. Even more selectionladen pictures of cultural evolution, which see cumulative cultural adaptations as the relevant explanandum, may not necessarily be associated with social learning mechanisms providing extremely high fidelity (like AM) or the accurate transmission of low-level details of traits (as shown by Henrich and Boyd discussion). To see where the RA's insistence on high fidelity qua AM comes from, we need to look in another direction. This last section explores the idea that the imitation-cumulative culture link is an analogy of the biological argument for Weismannian inheritance (Blackmore 1999).

A preliminary caveat. In the previous section, I said that the reaction against high fidelity replication in inheritance has become widespread, and its necessity to evolution by natural selection is seriously questioned. However, the same sources routinely admit that, while unnecessary for *general* evolution, replica*tion* (if not replica*tors*) might indeed be required for particular types of evolutionary products, such as *complex* cumulative adaptations and (on some accounts) evolutionary novelties (Godfrey Smith 2009; Bourrat 2014a,b). So complexity makes its comeback here. The core problem of this section will be to clarify why and how high fidelity is deemed central to the evolution of cumulative *complex* traits and adaptations. The explanation will crucially involve the idea that complex systems are *fragile* and that the cost of error is high. But let's start from the beginning.

To answer the above-mentioned question we must first go back to the model of genetic inheritance based on replication and to the genotype/phenotype distinction. As noted before, biological inheritance is often explained as the passing on of *genetic information* from parents to offspring. The question of how we should think of genetic information and more generally information in biology is notoriously a vexed one; nonetheless, it is fair to say that a common and very successful way of thinking of genetic information is as of a set of instructions, a recipe, a computer program and sometimes even a blueprint for specifying the phenotype. Such metaphors are essentially inadequate and simplistic, particularly when they are taken to mean that genotypes *determine* phenotypes, which entirely disregards the importance of development. Nevertheless, they are still central to the evolutionary debate. The idea of the genotype as a set of instructions or a recipe to "build" the phenotype leads to a view of inheritance that may be represented as follows, which I will call "Weismannian" after Dawkins:



Figure 16. From genotype to phenotype (adapted from Bergstrom & Rosvall 2008).

The horizontal arrows between G (genotypes) and Ph (phenotypes) represent the process whereby the "instructions" contained in the genome get "read" so as to construct the phenotype (with inputs from E, the environment, and D, developmental factors as well). The vertical red arrows represent the flow of hereditary information implemented by genetic replication, and are the *only* inheritance route available. According to the Weismannian view of inheritance, in fact, there is *no* flow of hereditary information from phenotype to phenotype (which would be Lamarckian inheritance, and would be represented by arrows connecting Ph₁ to Ph₂ and so on): this is prevented by what is known as "Weismann's barrier"— that is, the fact that the "instructions" provided by the genotype can only be "read" in the phenotype, but not "read back" into the genotype and then copied down to the next generation. In other words, any change intervening in the phenotype cannot work back to affect its generating matrix (the genotype), which Weismann gorily demonstrated by showing that mice whose tails had been cut off did not produce tailless mice.

Of course, extreme high fidelity transmission sits in this scheme as the red arrow representing genetic replication, which, as we said, occurs with a negligible rate of error *as mismatching*. The standard explanation for why G_2 needs to be extremely similar to G_1 is that any error *as mismatching* in these genetic "instructions" will lead to errors *as malfunctioning* in the expression of instructions into the phenotype—and, therefore, to errors (again as malfunctioning or even lack of functioning) in the organism's phenotype—as teratogenic mutations vividly show. Malfunctionings in the phenotype are important, because organisms are ultimately the targets of selection: organisms that don't "function" properly will be penalised or wiped out.

A very clear and playful illustration of this concept is provided by Ridley (2000). Imagine a group of children playing Chinese Whispers—a game commonly used to illustrate inheritance dynamics. The message passed on along the transmission chain is "WHEN THE TIGER COMES, FREEZE". Here there is an interesting tweak to the usual Chinese Whispers metaphor: every time a child passes on the message, a

228

bloodthirsty tiger *actually* enters the scene; in addition, it is assumed that the child receiving the message acts accordingly. Now the message is the analogue for genetic instruction; the child's actions, based on the received message, the analogue for the corresponding phenotype; and the tiger, of course, for natural selection. What happens is that every now and then one child misunderstands the message, hearing, for example "WHEN THE TIGER COMES, SNEEZE". The consequences are easily imaginable, and alas dire for the imprecise child. The gist of the analogy is that relatively small and trivial changes in the instructions (only two letters changed from "FREEZE" to "SNEEZE") can lead to disastrous consequences: there is a sense of *amplification* of one petty error into a much more serious one.

It is easy at this point (although perhaps not very convincing!) to imagine how the RA can be interpreted as the cultural counterpart of this reasoning, with AM somehow taking the role of fine-grained instructions to build the fully fledged cultural trait. But before I proceed to lay down the application of the idea to culture in detail (with its already evident problems), let me briefly look at the broader consequences of this line of reasoning on evolutionary orthodoxies (biological first, but then often generalised to abstract evolutionary thinking) and consider some final objections.

The consequences of the ideas just examined are numerous and scattered across different thematic areas. A very obvious one is what may be called the "gradualist thesis", which has been in fact identified as the biological version of the RA by Morin (2015). Gradualism essentially postulates that evolutionary change must occur in tiny sequential steps, rather than through spectacular "saltations", for small changes are less likely to be deleterious that large ones (Dawkins 1986). The requirement is particularly persuasive under the assumptions that biological mutations are "blind", i.e. occur randomly, and that the genotype-to-phenotype mapping is generally smooth (i.e. small variations in genotype lead to small variations in phenotype, Alberch 1991). Under these assumptions (which may not hold, however, in the cultural domain), taking a small step in an unknown direction is likely to prove less risky than taking a large one. Reflections of gradualism are

pervasively found in evolutionary theorizing—for example, we may spot its traces in lineage explanations (Calcott 2009) or in generative entrenchment theories (Wimsatt & Schank 2004; Wimsatt & Griesemer 2007).

However, one may still have reservations against gradualist ideas and requirements (Morin 2015), inasmuch as they depend on a genocentric picture of inheritance. Let us go back to the diagram of Weismannian inheritance. As mentioned in the previous section, this take on inheritance has been increasingly questioned by the growing body of literature on multi-systemic inheritance mentioned before, which convincingly shows that channels of inheritance from phenotype-to-phenotype do in fact occur (Jablonka & Lamb 2005; Griffiths & Gray 2001); in reality, most cultural inheritance seems indeed a case of phenotype-to-phenotype transmission, and even that the Weismannian barrier is not impenetrable (see for example methylation, Griffiths & Gray 2001; Bergstrom & Rosvall 2011).



Figure 17. Phenotype to phenotype inheritance may be less faithful, but is sufficient for evolution by natural selection .

The blue arrows connecting phenotype to phenotype are now representing a direct (some would say Lamarckian) form of inheritance between phenotypes—one that in effect, I think, represents the best model for most cultural transmission. One natural question at this point is: why bother evolving the complex machinery of genotype-to-phenotype mapping, with all the accompanying restrictions and requirements? Wouldn't it be simpler to just make phenotypic inheritance more faithful? From the previous section, in fact, we should also recall that nothing in the classic "recipe" accounts of evolution seems to require intrinsically the phenotype-genotype distinction. Such complex machinery seems to require a justification, and perhaps most importantly, so does our assumption that it embodies a universal evolutionary principle, to be generalised to other fields—such as culture.

This justification is actually rather difficult to retrieve in evolutionary theory; perhaps the best suggestion comes from a relatively recent piece of work by Bergstrom & Rosvall (2011), who discuss what they christen the "transmission sense" of information. What Bergstrom and Rosvall essentially point out is that direct and faithful transmission of complex developed structures like phenotypes is technically, empirically unfeasible. Faithful phenotype-to-phenotype transmission appears to be a pragmatically insurmountable obstacle. Why that would be the case is not fully outlined in their paper; what instead emerges more clearly is why, on the contrary, genetic material would be exquisitely designed for the task of storing, transmitting and retrieving large bodies of information, and thus an inevitable result of increasingly complex evolution. Their explanations point to engineering-like features of codes (of which DNA is a prime example) that make information stored in code-form better for faithful transmission and preservation than information in "uncoded" form (whatever that is!). Codes, they argue, offer a number of attractive (indeed, crucial) features to optimise transmission; these they describe variously as degrees of freedom for structuring redundancy (obtained through arbitrariness), possibility of compression while minimizing distortion, and compactness.

If Bergstrom and Rosvall are right, then, genetic material is the best (or the only?) solution to efficiently "pack" large quantities of information for transport.

Similar suggestions appear to come from Dawkins when discussing instruction-based Weismannian inheritance, which he describes as self-normalizing, idealised, and selfcorrecting (Blackmore 1999). Dawkins illustrates the concept through the example of instructions to fold an origami:

In the origami experiment, what passes to the next generation is not the paper phenotype but a set of instructions for making it. Imperfections in the execution of the instructions result in imperfect junks (phenotypes) but they are not passed on to future generations: they are non-memetic. ... This is because they all make reference to idealised tasks like "fold the four corners exactly into the middle". If the paper is not exactly square, or if a child folds ineptly so that, say, the first corner overshoots the middle and the fourth corner undershoots it, the junk that results will be inelegant. But the next child in the line will not copy the error, for she will assume that her instructor intended to fold all four corners into the exact centre of a perfect square. The instructions are self-normalising. The code is error-correcting. Plato would enjoy it: what passes down the line is an ideal essence of junk, of which each actual junk is an imperfect approximation. (ibid.:. XII)

Ultimately, then, the justification for the existence of code-like, transmission-friendly structures appears to derive from the necessity of packaging and compressing the large quantities of information necessary to build complex evolutionary products for efficient transport. In turn, it is essential that such "transport" happens highly conservatively (i.e. faithfully), as it does, because the "packaged" information will need to be re-expressed and read out into phenotypical entities—a process fraught with uncertainties and fragilities. ⁵²

⁵² There might be a further question connected to this reasoning, which is how did complexity come to evolve in a first place. If complexity depends on faithful transmission, but faithful transmission depends on complexity (because the genetic code is certainly a complex evolutionary product in and of itself), how did this complex circular project first take flight? The problem is known as the "catch 22

4.8. Conclusion

The discussion of fidelity has led me to develop various and rather eclectic threads, sometimes seemingly exceeding my declared objective: the disambiguation of the RA. In my defence, let me say that the question of fidelity's role in cumulative culture explanation is genuinely complex and at least as much under-explored. A thorough analysis of the topic would have required a dissertation on its own; a summary of the crucial issues shorter than the one offered here would have left out too much information to make the discussion intelligible. In effect, I think that the level ambiguity is not the only source of confusion in the conceptualisation of fidelity in the cultural domain: an equally important source of confusion is represented by the *content* of fidelity. When the information that can be culturally reproduced is multidimensional, like it clearly is in behaviour, what aspects of information are important to copy? What kind of similarity makes a cultural trait the "offspring" of another—what makes us identify, say, a lineage of hammer-like tools? Is it similarity of visual or functional features? These questions are important, but ,as I said, they shall remain unanswered here (and hopefully become food for future thoughts).

The take home message from this chapter is, of course, the identification of different usages of the term "fidelity" in RA doppelgängers. Reiterating the conclusions of section 4.4.3, we now have three different referents connected to the term "fidelity":

of life" and is supposedly solved by a "delicate process of bootstrapping" between complexity and fidelity (Godfrey-Smith 2009; Ridley 2000).

Type of fidelity	Level	Preservation, transmission, similarity of
Fidelity I	Sub-individual	Sub-components of traits (such as sub-actions morphology in AM)
Fidelity II	Individual	Whole traits in individuals (TP)
Fidelity III	Populational	Trait distributions in populations (also sometimes interpreted as TP at a collective level)

Table 3: Fidelity is conceived at three different levels

What is important about these notions of fidelity is that, despite all being derived from a reasonably stable meaning—i.e. the measurement of intergenerational similarity—they are not obviously coextensive nor easily correlated across levels of magnitude. In fact, I suggest that we treat them as straightforwardly incommensurable: if it can be established that premises of a line of reasoning contain different notions of fidelity, like many RA readings seem to suggest, we need an explicit account of how they correlate.

CHAPTER 5. Conclusion

5.1. Introduction

It is finally time to pull together the threads developed across this dissertation. The principal conclusion of this work is that the RA argument is only on the surface a single and univocal argument. In reality, it allows for a wide variety of interpretations, which depend on the different usage of its terms and on the different contents consequently assigned to the theoretical links L1 and L2 (and to L3, whenever this is drawn as a conclusion of the argument). In this chapter, I shall flesh out the most interesting RA interpretations and offer a brief evaluation of each. The focus will be on showing that these interpretations are genuinely distinct arguments, not reducible to each other.

Emphasizing this point is important, I think, not only because it demonstrates that philosophical analysis can uncover authentic ambiguities and thus contribute to a better understanding of arguments and models in cultural evolution studies (and perhaps help with further formulation); it is also instrumentally important to cultural evolution studies, for, I will argue, these different RAs are characterized by varying applicability in time, scope or phylogenetic breadth. Besides assigning divergent contents to the theoretical links, RA interpretations are chiefly distinguished by their explananda—they effectively target different aspects of the phenomenon that goes under the name of "cumulative culture". This result carries a double message. First, it suggests that "the RA" cannot be evaluated fairly as a single argument. Instead, each newly identified RA should be tested in relation to its appropriate domain, where it may be judged more or less effective in generating predictions, offering
novel insights, interpreting or organizing data. In other words, perhaps no RA convincingly explains "cumulative culture", but *some* RA may offer interesting insights into particular aspects of the overall problem. Second, the message is broader and forward-looking. The RA's multiplicity suggests that it may simply not be possible or productive to analyse "cumulative culture" as a homogeneous, definite occurrence. Instead, we need to develop, articulate and recognize a variety of conceptual tools, and be prepared to tackle the issue in a subtler, piecemeal fashion.

I promised in the first chapter that I would eventually pick up again the socalled Abacus diagram after filling the term "beads" throughout the chapters, and illustrate how different horizontal combinations would result in different lines of reasoning or arguments. However, the central chapters ended up identifying up to six different interpretations for the term "imitation", three for "fidelity", and at least four for "cumulative culture", so if one were to expound *all* the logical permutations of these numbers in the diagram, one would wind up with a bewildering kaleidoscope of 72 different ratchet arguments! Fear not: I do not illustrate, let alone examine, every possible combination here. The reason (besides obvious impracticality) is that not all of these interpretations are interesting and/or held in the actual debates. I will focus instead on the most significant ones, while trying to connect them back to their specific background and applicative scope. For visual reference, the Abacus diagram is reproduced in its "filled" version in the following section.

5.1.1. Terminology recap and filling the Abacus diagram

Before I delve into the last chapter proper, it might be useful to concisely recap the terminology previously obtained—and specifically the distinctions that will make it into the Abacus diagram. I do this very briefly in this introductory subsection, along with providing the final representation of the Abacus diagram. The first and obvious novelty in the new diagram is that the boxes corresponding to different accounts of

the three RA terms now appear filled. In addition, the boxes are connected by several coloured arrows. Same-coloured arrows intend to represent schematically the distinct logical paths that characterise each RA interpretation. A brief explanation for this notation and its use is provided in caption; however, I recommend coming back to the Abacus diagram only *after* having worked through the final chapter, rather than attempting to extract conclusive information about different RA interpretations from it. In fact, the diagram is best thought of as a thumbnail sketch of the results illustrated in the course of this last chapter. The information it presents is condensed to absolve this synoptic function, but may appear unintelligible without a suitably expansive explicative apparatus.

As for the terms distinctions, six different accounts of imitation will be represented in the Abacus diagram. The principal distinction is between generic imitation, or "imitation as all copying", and specific imitation, which I label "standard imitation". Standard imitation portrays a specific kind of copying rather than a general mimetic effect, which sources information directly from behaviour and is oriented to the reproduction of (nondescript) details of such behaviour. This account of imitation can be (and usually is) segmented further within the social learning literature, according to various criteria (see section 2.7): we thus obtain mimicry, overimitation, program-level imitation and mentalised imitation. All of these imitations, as standard imitation, may be read as detail-oriented types of imitation; the kind of details they pick on, however, make quite a difference to their significance (see 5.2).

Four general accounts of cumulative culture will figure in the Abacus diagram. As per Chapter 3, cumulative culture accounts derive principally from intersecting the qualitative vs. quantitative complexity distinction with the level distinction. This generates the four categories, i.e. qualitative/individual; quantitative/individual; qualitative/collective; quantitative/collective. For brevity reasons and to facilitate understanding, however, each account is represented in the diagram by a particular instance of concrete cumulative culture drawn from the corresponding category (rather than by the abstract category). We thus have: the technological trait (from

the qualitative/individual category); (from the the non-tech trait quantitative/individual category); the combinatorial repertoire (from the qualitative/collective generalist category); the repertoire (from the quantitative/collective category). Each account represents a favourite explanandum of a RA interpretation.

Finally, fidelity counts three competing interpretations, corresponding to the levels at which cultural transmission is thought to happen (and similarity between parental and offspring generations of cultural entities measured). As per the conclusions of Chapter 4, the three accounts of fidelity are: fidelity-I, corresponding to the transmission of sub-individual level details of cultural traits; fidelity-II, corresponding to the transmission of whole, individual cultural traits; fidelity-III, corresponding to the transmission of cultural traits collectives, traits distributions, or cultural repertoires.



Figure 18. Some possible interpretations of the RA on the Abacus Diagram.

(Continues from fig. 18.) As in the original Abacus Diagram (see Chapter 1 conclusion, 1.7), the standard outline of the RA is represented on top, in the first line of block-coloured square, circular, and fat arrow-shaped boxes.

The downward solid lines unfolding from the term boxes (imitation, high fidelity, cumulative culture) string together a series of dotted boxes, which represent the different accounts of each term isolated in chapters 2-4.

The space below L1 and L2 (fat arrow-shaped boxes on top) is occupied by coloured arrows connecting the term boxes, which represent different readings of the corresponding theoretical link for each RA interpretation. Thin arrows come in four different colours; each colour represents a different RA interpretation - as per the legend on the bottom right corner.

Purple arrows illustrate the Gradualist Interpretation. Standard imitation (which may be interpreted in turn as any of the more finely distinguished types of imitation, i.e. mentalised, over-imitation, program-level, mimicry) is necessary to generate fidelity-I, which in turn allows the preservation and accumulation of functional tech traits directly, or through fidelity-II (see 5.2).

Green arrows illustrate the Social Norm Interpretation. Standard imitation, or again some of the more specific imitation types (and especially mimicry and overimitation), is necessary to fidelity-I and consequently to the detailed preservation of the arbitrary form of the non-tech trait.

Yellow arrows illustrate various Quantitative Interpretations. Here, imitation must be assumed to be a very general mimetic mechanism, corresponding to social learning in general, or "all copying". Such imitation allows the preservation of entire traits (fidelity-I), or even of trait collectives fidelity-I), depending on interpretations. In these interpretations, the theoretical links L1 and L2 have nearly equivalent contents.

Magenta arrows illustrate the Exaptive Interpretation. Here, lines on the left side exceed the scope of the RA conceptual space, representing hypothesized connections between different types of imitation. In particular, mimicry is thought to enhance more sophisticated forms of imitation (see 5.2.1). In turn these allow the preservation of various type of trait details (fidelity-I), which contribute to the preservation of functional traits (fidelity-II) as in the Gradualist reading.

5.2. The Gradualist Interpretation

The first RA interpretation I shall present is what I've called the "gradualist" interpretation. Chapter 4 has already amply dealt with the logic that underpins it, so I sketch it briefly here and focus on divergences from other interpretations (particularly from what I call the "social norm" interpretation, described in the next section).

To understand where the gradualist interpretation comes from, we need to start from its target notion of cumulative culture, which is represented paradigmatically by individual technological artefacts and skills. Essentially, this interpretation treats tech entities as "cultural organisms", whose complex internal organization and connected "viability" can be disrupted by relatively minor changes. Fidelity is hence extremely important, because fidelity is responsible for controlling the amount and amplitude of such changes during transmission. But which fidelity?

The gradualist interpretation is undoubtedly characterized by its insistence on type-I fidelity, i.e. the transmission, inheritance or similarity of sub-components of individual traits. Type-I fidelity is the notion plugged into L1 and thought to be supported by imitation. In gradualist interpretations, imitation is some kind of social learning mechanism that has high-resolution, near-digital qualities, being sensitive to details and subcomponents of action sequences. What exactly these "details" or "subcomponents" are is far from clear: are they topographical details of actions (limb-to-task selection, amplitude of movement, etc.)? Then the relevant kinds of imitation may be overimitation or mimicry. Or are they causal features of actions and their temporal structure (things like the casual role of a handle relative to a percussive head, or the preparatory role of one action relative to another in a sequence, etc.)? Then the relevant kind of imitation might be program-level imitation. Or are they intentional features of sub-actions (such as the goal of a movement, even if unrealized; the choosing of a particular shape and kind of material to a particular end)? Then the relevant kind of imitation may be mentalised imitation.

While the potential of these different kinds of imitation is far from irrelevant to the argument's applicability, all of them are amenable to a description in terms of "detail-oriented" imitation, thus supporting the gradualist reading of the argument. All these types of imitation are broadly compatible with what I have called "standard imitation" (cf. 5.1.1). Therefore, for the purpose of illustrating concisely a range of gradualist interpretations, which appear to lean towards specific subtypes of detailoriented imitation (but often implicitly), I have represented "standard imitation" as logically central in the Abacus diagram. In gradualist accounts, the importance of standard imitation kinds to type-I fidelity varies from being necessary to highly conducive.

What sharply distinguishes the gradualist interpretation from others is the hypothesized content for L2. Why and how is type-I fidelity necessary (the necessity-based reading is virtually universal) to cumulative culture in the form of technological artefacts and skills? For technical skills (things like adhesive preparation, throwing proficiency, processing of toxic plants, etc.), the answer is quite straightforward: the reproduction of details of such behavioural sequences is essential in one way or another to the correct acquisition of skill, i.e. to the acquisition of a skill that achieves the promised benefit (successfully attaching a spearhead to a shaft, hitting the target reliably, detoxifying plant food, etc.). Failing to reproduce one or more steps will result in hampered transmission of the skill and, in all likelihood, a fitness-reducing disaster.

The reasoning for artefacts is the same; but because artefacts may be construed as material products, which persist after the particular behavioural process that generates them, they are even more apt to illustrate the gradualist logic. This assumes that the artefact and its correct functioning depend univocally on the generative procedure or behavioural recipe that imitation faithfully transmits much like an organism's viability and fitness depend on the errorless transmission of its genetic code. This point is ultimately an empirical question; however, it is widely assumed to hold, for example in the literature about lithic technologies (Hiscock 2014). As I noted in chapter 4, such logic retraces the biological principle of Weismannian inheritance in biology, shifting the requirement of fidelity to a generative matrix rather than to the final product: the correct form of an artefact depends on the faithful transmission of a set of instructions required to build it. Error-as-mismatching in the instructions or "recipe" translates into (and often amplifies) error-as-malfunctioning in the final product; error-as-malfunctioning results in counter-selection, i.e. the trait being abandoned and/or failing to be retransmitted (recall the unfortunate children facing the blood-thirsty tiger in 4.7, whereby minor transmission mistakes resulted in disaster). The more complex the artefact, the longer and more error-prone its generative recipe; the higher the chances of failure. So type-I fidelity becomes imperative. Thus the gradualist interpretation assigns a very specific content to L2: *fidelity is necessary for the evolution of cumulative culture because technological traits, and perhaps artefacts in particular, are complex and fragile entities.*

Interpreted in this manner, the gradualist reading also provides a clear and explicit way of bridging the gap between type-I and type-II fidelity, explaining how and why fidelity-I is important to fidelity-II, the preservation of the entire trait. Fidelity-I makes it more probable that the entire trait will preserve its functionality; preserving functionality is indispensable for the survival of the trait. (Indeed, I suspect that function has a lot to do with the identity of technological traits.) This is at heart a selection-based explanation: by interpreting the functional dimension as crucial to the maintenance of the trait, gradualist analyses uncover a selective facet to the problem of transmission. A trait that superficially resembles a hammer morphologically, but does not work as a hammer, is likely to be selected against and fail to spread further (unless it acquires a different function—say, that of a toy). Many of the traits targeted by gradualist interpretations of the RA are in effect commonly depicted as "cultural adaptations", travelling through history via incremental, selectively shaped modifications (Boyd et al. 2013).

So far, my exposition of the gradualist interpretation has been mostly descriptive—I have not devoted many words to its evaluation. Let me now try to give

a general assessment. As I have anticipated in chapter 4, there may be problems with it, but also elements that make it one of the most interesting and justified interpretations of the RA. Difficulties concern, in particular, those gradualist arguments that assume a somewhat "dumb" definition of imitation as mere action matching, such as mimicry and overimitation, for example. The rest of the discussion takes therefore into consideration gradualist arguments that interpret imitation in this way—which are, I think, the most popular.

The gradualist/Weismannian interpretation of L2 hinges on two controversial empirical assumptions. One, as noted, is that the final trait depends on a rigid and fixed motor procedure, a "behavioural recipe". The second is that cultural change is blind—as genetic mutation is (this is necessary in order for the Weismannian analogy to hold).

These assumptions are often contradicted by empirical evidence. First, in many cases the final trait does *not* seem to depend univocally from the performance of a rigidly determined sequence of movements. In section 4.4.2 I mentioned many cases in which action-matching can hamper the acquisition of a trait rather than helping it. It is quite obvious that understanding the causal and procedural nodes is often far more important to successful trait reproduction than the dull replication of fine topographical details—all the while providing a certain degree of freedom in execution. Second, if causal understanding is the real deal, it is likely to produce something more similar to guided than blind cultural variation. In many cases, agents can readily see and tell whether a modification is a mistake or not: they are clearly not groping in darkness; they will steer clear of fruitless modifications, aiming for useful one.

Nonetheless, there may be a limited but important subset of cases where pedantic reproduction is in fact a useful or obligate strategy—bringing about something akin to blind variation. These would be cases of technologies that are error-intolerant, on one hand, and, on the other hand, causally/procedurally opaque, counterintuitive, and hard to reverse-engineer by a single individual.

In this context, "error-intolerant" means fragile and dependent on a univocal (or nearly univocal) behavioural recipe. Acerbi and co-authors, for example, identify various occurrences of traits displaying what they call a "rugged landscape" of possible modifications to their generative matrix (the behavioural recipe), causing abrupt drops in efficiency correlated with variation (see Acerbi, Tennie & Nunn 2011). These sort of skills or technological apparatuses are more likely to depend on pedantic copying, because mistakes might mean total breakdown, rather than moderately deficient, improvable performance.

Counterintuitive, causally opaque, and hard to reverse-engineer features indicate instead a particular complexity of the task. The complexity is structural, but above all cognitive: it is simply hard to figure out *why* that particular piece goes in that particular position, in that particular stage of construction; *why* this ingredient needs to be added before and not after the other; *why* the shape of a throwing implement improves its trajectory; and so forth. Importantly, cognitive complexity is a *relative* property: it depends not only on objective features of the task at hand, but also and crucially on the cognitive sophistication of the performer. Thus, the more the agent's cognitive penetration is limited and causal understanding poor, the more a trait may qualify as causally opaque, hard to reconstruct, etc.

In these cases, variation may be legitimately construed as blind. Not only may successful technological innovations be the result of "lucky accidents" that get incorporated into the procedure without a proper understanding of their workings (merely because *they work*)—but it also becomes imperative that further occurrences do not depart excessively from the serendipitously achieved structure, because the directions of improvement are at least as impenetrable as its "mysterious" working.

The crucial question then becomes: how frequent and/or significant are such cases, and are they common enough to justify the gradualist RA?

I think the picture sketched above may plausibly apply to a variety of *ancient* but sophisticated adaptive technologies: examples may range over such things as bow construction, boomerang shaping and throwing, adhesives concoctions, and

specialized tool, shelter or clothes making. These activities are complex and hard to devise even for modern minds; it seems sensible to hypothesize that they would have been even more so for proto-hominins, whose cognitive power was plausibly limited and causal understanding correspondingly deficient. In many of these cases, the reason that a technique works will be obscure not only to the novice learner but also to the skilled performer—who might have stumbled on the efficient solution by chance, subsequently maintaining it. Contemporary ethnographic evidence appears to confirm that in some cases justification for a particular procedure is lost to the understanding of the entire pool of its performers (Henrich 2015). In addition, the archaeological record often shows extreme conservatism and "vestigial" maintenance of suboptimal parts for these technologies (cf. Balfour in Chapter 3; Sterelny 2003); both observations being compatible with the hypothesis of "fumbling" improvements.

The reach of the gradualist RA may therefore be chronologically limited to deep-past technologies, perhaps confined to preverbal or proto-verbal times, and perhaps particularly to those that can be mastered by single individuals or very small groups (again, to fully exploit the explanatory potential of basic imitation). This may be a narrow reach indeed, but not an insignificant one: deep-past technologies are among the factors that bootstrapped humanity through cycles of territorial expansion and resources exploitation, which in turn played a crucial role in the evolution of our species (Henrich 2015; StereIny 2016). A fruitful direction for future research could therefore be to concentrate on compiling a list of concrete and documented ancestral tasks that are simultaneously error-intolerant and cognitively challenging.

5.2.1. Adding depth to imitation: a speculative tweak to the gradualist interpretation, or the Exaptive Interpretation

I will conclude the discussion of gradualist interpretations by sketching a couple of potential objections—and responses—to their explanatory reach.

The first objection is optimistic (it aspires to expand the reach of the argument) and concerns the temporal extension of the argument. It seems that many modern technologies are also incredibly hard to reverse-engineer in the absence of specialized knowledge: think about recomposing a smart phone after a ruinous tumble, or assembling a piece of furniture without instructions. Perhaps standard imitation has something to do with the increasing complexity and diffusion of these technologies too? I think the answer to this question is essentially negative (though see the "patent" interpretation in 5.4). If we want to stay true to the classic RA reading, and its designation of a relatively simple visuo-motor mechanism like imitation, modern technologies are not likely to be a plausible explanandum. The development and transmission of such traits is entirely conditional on a host of late interdependent adaptations-for instance, natural and mathematical languages, writing, intensive and global cooperation, scientific methods, etc. If standard imitation still has a part in this, it is likely a negligible one, and hard to disentangle with precision; in addition, it can hardly be modelled as a blind innovation-searching mechanism in such a context.

The second objection is pessimistic (it seeks to shrink the reach of the argument) and perhaps more damaging. It concerns not so much the temporal, but the *actual* explanatory reach of the gradualist interpretation: the idea is that the RA is simply too restrictive to be significant, fitting at best an insubstantial range of proto-technologies. Framing this objection and providing a reply will require taking a few steps back; at the same time, it will eventually allow me to sketch an alternative reading of the RA that preserves the intuition about an early role of simple imitation mechanisms but attempts to overcome their limitations by suggesting an ontogenetic link to more rewarding (and demanding) imitation types.

I have so far illustrated the case of gradualist RAs that assume the "simplest" types of imitation—those that can account merely for the faithful reproduction of motor details of behaviour. In essence, these arguments assume that the *motor*

organization of the task and its topographic outline are good enough proxies for the maintenance of its procedural, causal, and functional properties (since functionality is what matters to trait maintenance and diffusion). Whether this assumption is correct is ultimately an empirical question: it is entirely possible that, on closer investigation, the picture sketched is indeed too simplistic and narrow. A famous objection to the transmission potential of "dumb" imitation is raised by Heyes (2013): she questions the importance of the reproduction of actions and bodily movements (i.e. imitation) and argues in favour of what she calls "emulation", i.e. the reproduction of the relative positions and interactions of animate and inanimate bodies. This kind of information, which sometimes goes under the name of "affordances", may be indeed very important to the procedural understanding of a large number of tasks, particularly those involving close and intensive manipulation of components. Vital skills like throwing, ambushing, or endurance running would remain purely motor and susceptible to standard imitation—but a substantial portion of tool-crafting and tool-usage tasks would clearly benefit from affordance learning.

While Heyes' objection is sensible, I think it is uncharitable to the gradualist interpretation—or at least to the intuition that simple imitation mechanisms can play a central and interesting role in the evolution of cumulative technology. Earlier I said that standard imitation is compatible with various specific definitions of imitation, all characterizable as the reproduction of sub-individual details or aspects of entire cultural entities: this included mimicry, overimitation, program-level imitation and mentalised imitation. However, I discussed only gradualist interpretations that assume the purportedly "simplest" (or, as Heyes calls them, "dumb") definitions of imitation—i.e. mimicry and overimitation (according to the usual reading). In effect, only *these* readings are susceptible to Heyes' objection; other readings, for example adopting program-level or mentalised imitation, are beyond it. So why not defend gradualist readings of the RA based on more sophisticated brands of imitation?

The problem with sophisticated forms of imitation is twofold, I think. Most obviously, they run a risk of "diluting" gradualist interpretations (similar to the one mentioned for modern technologies): "fancy" forms of imitation are not good metaphors for blind variation. Surely, imitative phenomena like program-level and mentalised imitation are the best-placed candidates to provide the kind of broadspectrum, transversal, across-features fidelity relevant to the largest pool of tasks ancient tasks included. If you select a superior type of rock to knap, being able to reproduce only the topography of your actions (dumb imitations) may in fact not be an efficient guide to the right material for me (for I may be in a location lacking that source of material). By contrast, the capacity to copy texture, colour, and whereabouts of that rock would be a more reliable guide. The ability to copy these kinds of detail may depend on understanding your intentions—in turn suggesting that mentalised imitation is the mechanism responsible for this skill. Likewise, abstracting away from low-level features of action, mining and copying procedural information is likely to be a faster, more efficient, and more reliable route to functional acquisition of trait—something that would require program-level imitation.

However, the kind of high-level capacities underpinning these sophisticated imitation modes—intention reading and causal understanding—are not cheap and not early. If these capacities are fully in place, it is easy to imagine more independent, creative, and self-correcting agents; perhaps even linguistic agents, whose reliance on behavioural information extracted via imitation is supported by alternative and powerful routes of transmission. Eventually, plugging such notions of "imitation" into the RA downplays the significance of the imitative mechanism in favour of a broader skill set for transmission; all the while detracting from a properly gradualist reading, which hinges as we said on the assumption of blind variation. This is not to say that such forms of imitation could not play a plausible evolutionary role—on the contrary, they very likely did—however, their role is not well-captured by the gradualist metaphor as discussed here.

The second problem with plugging sophisticated forms of imitation into the argument is more substantial than formal, and concerns chronology. As I said, even if these kinds of imitation diminish the gradualist interpretation, they are clearly quite apt to cultural transmission. But they are late in emergence, raising the interrogative of what was going on before them.

I think that a plausible solution to these problems may come from attempting a *reshaping* of the RA. In effect, it is unclear whether what I am going to suggest can still be called a "RA interpretation", given that it essentially abandons the connection of L1 and L2, in favour of a prominent and heavily revamped L3. It has elements in common with gradualist RA interpretations in that it starts out from simple imitation mechanisms, but it works its way through a more indirect, diachronic and ultimately different kind of reasoning.

I will illustrate this solution very briefly, and admittedly speculatively. The central idea is that the debate over the "right" or "best" social learning mechanisms for the RA is framed in a misleading way: instead of representing competing alternatives, different types of imitations may stand in a relation of increasing ontogenetic and evolutionary complexity, and simpler imitative phenomena may facilitate more demanding ones through the establishment of dedicated neural circuitry.

To begin with, there are various reasons to be sceptical of the current framing of the imitation question, and to wish for a change in focus. First, it is unclear whether and to what extent the various imitation mechanisms examined in Chapter 2 are truly sharply distinguished, and how neatly they map the cognitive setup of real organisms. As I have noted, categorization efforts have often been inspired by incongruous motivations. Second, various lines of evidence suggest that the evolution of complex cognitive capacities is often based on a system of self-feeding loops (StereIny 2012). They do not pop out of a lucky genetic mutation, or of a single, punctual, lineage-changing event; rather, they emerge gradually from tinkering on simpler scenarios—they are themselves, in a sense, "cultural products" (Heyes 2016; 2018). It is not farfetched, then, to imagine a similar trajectory for complex imitation mechanisms, emerging haltingly and progressively from a host of simpler capacities, perhaps involving simpler imitations as well. Third and final, while there are apparently more and less complex forms of imitation, a "hierarchy of complexity" needs to be formulated with caution. Mechanisms such as program-level and mentalised imitation are evidently demanding, but it is less obvious that mimicry and overimitation, for example, can be characterized as "simpler" or "earlier" imitations. Overimitation, in particular, is only apparently a "dumb" or simple mechanism. The label derives from a puzzling abdication of individual rationality—most likely betraying *complex* and derived cognitive control—rather than from an appeal to elementary cognitive functions. If one looks at overimitation from the perspective of a putative common ancestor (often modelled on present-day chimpanzees), the results are revealing: chimpanzees show no evidence of spontaneous overimitation—they predominantly rely on fast and efficiency-maximizing heuristics like emulation (in the sense of copying the goal). Equally tellingly, chimpanzees show some (flimsy) evidence of privileging social information and copying "pedantically" only in a very limited set of circumstances, namely when exposed to causally opaque artificial tasks (Whiten et al. 2009). This result has been interpreted as suggesting that overimitation represents a culturally evolved capacity, reinforced by increasing reliance on complex cultural tasks, rather than a primitive and simpler capacity. As a matter of fact, I think, overimitation looks not entirely dissimilar from mentalised imitation, and definitely more like the end product of evolutionary tinkering rather than a step in the ladder.

My opinion can be summarized as follows. Within the human context, the reliable and incremental transmission of functional cultural traits owes much to sophisticated imitation types, and perhaps especially to program-level imitation first (chronologically speaking). Program-level imitation is demanding, but it has been described in nonhuman primates (Byrne & Russon 1998); it provides the kind of control and flexibility over technical actions that is likely to support good levels of functional maintenance and perhaps fruitful innovation, especially in fitness-impacting tech traits.

However, program-level imitation is not a primary type of imitation (and nor is mentalised imitation, for that matter); it is instead the by-product of simpler forms of imitation—of mimicry, perhaps.

The reason that it is hard to conceive of program-level imitation (or any other form of sophisticated imitation) as primary is that selective explanations of its emergence appear wanting. Forms of "fancy" imitation are good for cumulative technology, no doubt; but can we really say that this proficiency explains their emergence and distribution? I think the answer is no; these forms of imitation have not been selected for their contribution to cumulative technology. They are not, then, adaptations for this end-nor they could be. In fact, no imitation mechanism (not even contemporary mentalised imitation!) is a fool-proof guarantee of trait acquisition: in real settings, complex tasks are virtually always learned, I would say, via something that may be termed "trial-and-error imitation". This would typically look like a lengthy, diachronically spread series of attempts guided by social information but made efficacious by personal practice. The fitness-accruing result of this process is at the *utmost edge* of this experiential sequence; it is difficult to imagine how such a distal event (even when successful) could act as a positive reinforcement on the disposition to keep copying. The disposition to copy should have, it seems, a more proximate motivation—or piggyback on mechanisms that can allow such proximate motivations.

Here is where mimicry makes it entrance. If the accounts that associate mimicry with relatively simple phenomena (such as, for example, synchronous movement and contagion) are correct, then this form of imitation may represent a truly foundational capacity. The way I see it, with respect to information-processing, mimicry simply indicates that the minimal cognitive gear for observation-to-performance action is in place (in other words, the presence of a neural correlate solving the correspondence problem—be it mirror neurons or vertical intermodal matching). But it is the *motivational profile* of mimicry that makes it the most interesting candidate for the role of stepping stone in the imitation story. It seems that mimicry is both triggered and reinforced by affiliative, prosocial sentiments.

Thus, mimicry in very young children may be motivated by the positive response of adults—by the simple drive to be liked, with all of its pragmatic consequences in terms of survival chances. This makes the reinforcement mechanism for mimicry easy to ground in plausible evolutionary scenarios—the "repetition compulsion" becomes credible.

In turn, mimicry might have acted as a reinforcement to action *understanding*—which is presumably the crucial cognitive capacity underpinning program-level imitation. Reinforcement may be conceived of as both ontogenetic and phylogenetic. Young learners with a predisposition to mimic for social reasons may transfer such a predisposition to other domains (this may even depend on the role of neural pathways enhanced by previous imitative activity); the tendency to repeat actions could lead to a better understanding of the internal organisation of motor behaviour. The old saying *repetita iuvant* seems to be at the basis of, for example, martial arts drills. Pupils are made to repeat, repeat and again repeat the same movements over and over again; it is not so much instructions that will make the difference, but some sort of unfathomable motor knowledge derived from sheer reiteration.

Once good mimickers have gained themselves a place as better-than-average program-level-imitators, other "reinforcement" mechanisms may take place. (This kind of reinforcement, acting predominantly on the phylogenetic, intergenerational scale, may more properly go under the name of "selection".) Better-than-average technological imitators may start seeing the benefits of their competence; they may further pick among their mimicking offspring, selecting for higher baseline capacities, further amplifiable by an increasingly "encultured" environment. Other sophisticated imitations may emerge during this process, driven by new capacities and new selective environments.

Incidentally, this account suggests in its entirety a picture of imitation as an *exaptation* rather than an adaptation in our evolutionary history: imitation wasn't originally selected for its role in cultural evolution, but for some social

cohesion role; it *then* got refitted to the cultural role (cf. the RA accessory claim VI in section 1.6).

In its present form, this account is clearly closer to evolutionary storytelling than to an evidence-supported hypothesis. Nonetheless, I think it has elements of plausibility; above all, it has the merits of being an improvement on the excessively schematic picture of imitative mechanisms that may be plugged back into the RA, and of further exploring the potential significance of basic imitation mechanisms. The real story is of course likely to be more complex than this, and to involve far more actors than simply *imitation*; within the limited conceptual space of imitation hypotheses, however, it is worth remembering that "imitation" is a multifaceted and non-trivial phenomenon.

Now, let us look at a second interpretation of the RA: I call it the "social norm" interpretation.

5.3. The Social Norm Interpretation

What I call the "social norm" interpretation of the RA is in various ways the opposite of the gradualist interpretation. It assumes a similar reading of L1: something like standard imitation (and, even more markedly, mimicry or overimitation) is thought to provide type-I fidelity. But the two interpretations differ sharply in the content assigned to L2, and in the choice of the relevant cultural explanandum.

The form taken by cumulative culture in the social norm interpretation is that of an individual non-technological trait (which may plausibly multiply and diversify into a repertoire, but the explanatory focus is first on local, individual traits). I have mentioned a few examples of non-tech traits when discussing functions and efficiencies in chapter 3, to which I shall add a few more: ritual dances, such as the "pizzica tarantata" of Southern Italy, thought to help disperse the venom injected by spiders and causing various types of mental illness and/or social stigma; traditional codified greetings, like the Māori *hongi* (pressing simultaneously one's nose and forehead to another's in greeting); more broadly, items from the following categories: fashions and trends, such as the Japanese urban sub-culture of ganguru girls; pretty much all linguistic and symbolic conventions; "group membership badges", ways of walking, eating, speaking and managing social interactions that are idiosyncratic to a group.

All these things, I have argued, have "quirky" functions and efficiencies. It seems incorrect to claim that they have no function whatsoever: indeed, many of these traits may be suggestively re-described as "social technologies", for they result in some form of social manipulation, control or engineering (they may comfort, unite, stir, cure, inspire, etc.). Even then, however, they differ from literal technologies in critical ways. First, agents often appear to have a less clear understanding and awareness of their instrumental value, less or no command of their workings and outcomes, and diverse motivations for adopting them. Second, the instrumental impact of such cultural traits seems to depend only evanescently, if at all, on the material features of the trait—certainly not in the same way as the percussive force of a hammer depends linearly on the length of its handle. Indeed, the same physical features in a non-tech trait can produce different outcomes in different communities (think of words' sounds). Third, and relatedly, it is hard to devise or measure incremental efficiency in these traits, either locally or historically.

I have hypothesized that non-tech cultural traits owe their odd functionality (when present at all) to the gathering of human consensus and recognition around an *arbitrary* form of the trait. This makes function less anchored to the material, physical, perceptual features of the trait; and less firmly predictable and comprehensible than in literal technological cases, for human consensus shifts, scatters and converges quite erratically over space and time.

Why, then, would high-fidelity imitation be necessary or particularly conducive to the transmission of such traits? The answer to this question (ultimately providing the content of L2) is profoundly different from the one provided by the gradualist hypothesis. Fidelity is necessary because non-tech traits are *arbitrary* creations,

rather than complex or fragile ones. For tech traits, functional properties can in principle be understood independently of their contingent task structure, and transmission based on such understanding would arguably be even more reliable and progress-oriented; here pedantic copying is an obligate strategy responding to suboptimal procedural comprehension, effective only to the (limited) extent that topographical details of behaviour act as reliable proxies of causal and procedural information. Nothing like this obtains for non-tech cultural traits: their functional value cannot be inferred reliably from features of the trait; there is no alternative transmission route other than blindly copying the trait.

Consider the tarantism example mentioned above: "patients" of this spectacular form of choreutic therapy were thought to restore their health (and social credibility or acceptance) through "dancing" at the rhythm of a traditional instruments ensemble—always including at least one tambourine. The "dance" appeared like a trance-inducing frenzy, typically including dramatic arched-back movements, highly reminiscent of the classic symptoms of so-called hysteria (Daboo, 2010). These movements clearly had no actual therapeutic power, but they were nonetheless reliably reproduced, and often read by onlookers as tangible signs that the spider's venomous bite was being exorcised. One could read in these theatrical bowing movements a parody of a "venom expulsion act": but why not mimic a retching spasm instead? These movements were unnecessary and inefficient from any reasonable functional perspective, but, being socially *interpreted* as therapeutic, they had to be reproduced in the established form in order to achieve their purpose.

The same reasoning holds for non-tech traits that are even less clearly functional: fashions and trends owe their form and existence entirely to the dull reproduction by new actors. Again, fidelity is necessary here because non-tech traits are largely arbitrary entities, not because they are complex of fragile.

The sense of accumulation in this interpretation is different from the one prevailing in the gradualist interpretation: sub-components of non-tech traits are unlikely to be linked together by causal and normative interdependencies, which would require a specific order or layout in addition of details. Whether the pizzica

arched-back movements are followed or preceded by feet-stomping movements is irrelevant: arguably, the importance lies more in the visual similarity and recognisability of each movement. No doubt there may be more sophisticated and codified forms of ritual expression that require rigid observance of a specific sequence of movements, but the requirement, again, is imposed by the observers, rather than by an inherent normative organization influencing the outcome. Lenient observers, or prestigious performers, could in principle grant greater structural liberty; in technical tasks, by contrast, this is simply not an option. As a consequence, accumulation is additive rather than qualitative; it loses its gradualist, incremental quality.

Let us come to the evaluation stage. Is this interpretation of the RA legitimate and useful? I think it is. Heyes (2013) defends something very close to the social norm interpretation of the RA. The social transmission of non-tech traits via "dumb" imitative strategies may have had an important role in the evolution of our cooperative dispositions—for example, by enhancing group membership insignia and ensuing intra-group cohesion. In addition, this interpretation of the RA may still be significant to later cultural forms: perhaps current non-verbal social norms and insignia largely owe their spread and accumulation to imitative phenomena.

5.4. Quantitative Interpretations

I have spoken of "quantitative interpretations" sporadically in Chapters 4 and 3, mainly with reference to Avital and Jablonka's notion of "animal cumulative culture" and Lewis and Laland's account of fidelity as trait preservation. These accounts present what I call a "quantitative" reading of the RA because they construe cumulative dynamics as merely additive processes, something akin to the "piling up" of traits or modifications. Accordingly, they adopt an underlying notion of complexity that is purely quantitative—what I have called "Oswalt's complexity" (cf. section 3.6.1). I will provide a quick précis of these RA interpretations and discuss their weaknesses and potential merits.

The first quantitative interpretation that should be mentioned is the one connected with Avital and Jablonka's notion of animal "cumulative culture". As I remarked in chapter 3, this notion of "cumulative culture" is effectively indistinguishable from that of plain *culture* (which is notoriously broadly formulated, in evolutionary settings, to cover instances of nonhuman traditions). It is perhaps legitimate to construe "accumulation" in a purely quantitative sense and at the collective level of the repertoire' but it certainly dampens the sense of complexity inherent to "proper" cumulative culture notions, as well as the fundamental explanatory endeavour of the RA, reducing it to a mere re-description of regular and obligatory cultural dynamics (those without which no culture whatsoever would exist).

As for the other two terms in the argument, "fidelity" features as what I have called "type-II fidelity": trait preservation (or perhaps even fidelity-III, repertoire preservation, but certainly not as fidelity-I). The macaques' society described by Avital and Jablonka acquires first the potato-washing trait and maintains it; it then adds the water-dwelling trait and maintains it; then comes the turn of fish-eating, and so on. Imitation, we are forced to infer, must be some kind of generalist mechanism with mimetic effect (an outdated sense of imitation as "all-copying", i.e. all social learning). In fact, Avital and Jablonka's interpretation precludes any appeal to imitation as a specialized social learning mechanism—being that the macaques' learning style is generally characterized as stimulus or local enhancement. But this makes the explanatory role of "imitation" in L1 terribly weak, reinforcing the suspicion that the whole line of reasoning really just describes good old plain *culture.*⁵³

⁵³ A brief exegetic note on Avital and Jablonka's work. Here, I construe their reasoning as an occurrence of the RA on the grounds that it explicitly exhibits a competitive explanatory route to the same phenomenon: cumulative culture. In effect, however, the authors originally aimed at *dismissing* the RA, showing that "cumulative culture" can be obtained *in the absence* of any specialized learning

A slightly more qualified version (at least in intention) of the quantitative interpretation is the one associated with Lewis and Laland's account, analysed in section 4.2.2. Recall that these authors present fidelity as "the key" to cumulative cultural phenomena, seeing their work as a formal ratification of verbal RAs. They provide an explicit definition of "fidelity" as *trait preservation* (the reverse of trait loss), which translates as type-II fidelity in my level-sensitive framework. While it is certainly true that trait preservation is necessary for cumulative culture, this reading of L2 seems again to debunk the argument as being little more than an ambitious redescription of plain cultural phenomena. Except for its having incorporated human examples of traits in its models ("driving a car", "speaking a language" are some of the examples offered), one struggles to see the difference with Avital and Jablonka's proposal.

The account presents a parallel drawback when it comes to imitation and the content of L1. As we are offered a definite and explicit account of fidelity, it seems natural to apply it consistently throughout L1 and L2. But, then, how would imitation be especially suited to provide trait preservation in L1? We again need to downgrade imitation to a generalist mimetic phenomenon (all-copying), or to simply assume within L1 an undisclosed account of how this particular social learning mechanism delivers trait preservation better than others. If this undisclosed reasoning is the customary one—namely, that something like *standard* imitation provides fidelity as action matching in L1—then fidelity must be fidelity-I. But this would burden the argument with inconsistency about levels of fidelity and demand a credible explanation for filling the gap between type-I and type-II fidelity (as discussed in

mechanism such as imitation. Avital and Jablonka's argument may ultimately be considered a "bogus RA": imitation (in the usual lectio of specialized learning mechanism) never had a place in it to begin with. Nonetheless, I hope to show with this slightly artificial reconstruction that quantitative interpretations along these lines have had a tangible presence in the literature. A clear example of this, flaunted as subtler but not so different in substance, is Lewis and Laland's account presented below.

chapter 4). In sum, under a charitable reading—i.e. reducing ambiguity for fidelity this quantitative version of the RA borders on triviality, being a mere re-description, rather than explanation, of plain cultural dynamics, and remains wanting on the side of L1 (failing to characterize imitation or motivate its presence in the argument).

But what if, instead, we give Lewis and Laland a *less* charitable reading? Perhaps the authors don't have in mind anything like fidelity-I in L1 and fidelity-II in L2; perhaps they simply don't register level discrepancy because they think the gap between fidelities is irrelevant—or better: *non-existent*, a false problem. If fidelity-I correlates linearly with fidelity-II, in fact, the gap is automatically filled, and no explanation is required. As it turns out, I think that Lewis and Laland's account implies exactly this kind of misguided reasoning.

Let me say first why I think the reasoning is misguided. As I hope to have sufficiently demonstrated in chapter 4, the gap between fidelities at different levels is real. In the actual world (i.e., beyond mathematical modelling) preserving details of traits doesn't automatically guarantee the preservation of entire traits (see 4.4.2); nor is it true that low-fidelity at the level of individual traits spontaneously translates into low-fidelity at the level of the cultural collective (see 4.4.3). Why this is the case is a theoretically challenging matter that I cannot hope to adequately address here: perhaps some of the relevant justifications may be construed in terms of emergent properties of complexity, or holistic organization of cultural traits. The fact that a phenomenon is not reducible to the arithmetical sum of its parts is bound to disrupt the logic of "digital" transmission: transmitting parts will not secure transmission of the entire (holistic) phenomenon, by definition. Or it may be that we deploy a far too simplistic and uni-dimensional picture of the flow of information in biological and cultural systems. In any case, as I said, establishing the causes for this state of affairs is beyond the scope of my analysis here: what matters, and what I will reiterate, is that fidelities are *incommensurable* across levels. By this statement I mean that fidelities cannot be placed on a uni-dimensional "digital" gradient, where switching from one level to the other is merely a matter of zooming in or out. Sometimes discussions present fidelity-I as simply "more accurate" than fidelity-II, and the latter

as "more accurate" than fidelity-III—as if switching from one to the other were merely a matter of resolution. This metaphor may be correct in some cases, but we should not assume it holds in *every* case. Indeed, it would be ideal to exercise caution first, and start from the assumption that fidelities in cultural transmission are in principle incommensurable.

But let us go back to Lewis and Laland's account. On one hand, I claimed, this appears essentially equivalent to Jablonka and Avitals'. On the other hand, however, it lends itself to a different and (at least apparently) more sophisticated reading. Recall that Jablonka's notion of cumulative culture takes the entire repertoire as the object of "complexification" and cumulative dynamics: it is an example of quantitative-*collective* notion of cumulative culture (see 3.9). Instead, Lewis and Laland's is a quantitative-*individual* account of cumulative culture, where the accumulation of modifications happens at the level of single traits first.

This is evident when we look at the way "events" are construed in Lewis and Laland's models. Recall that they have target populations picking cultural traits from an external finite pool of ten "seed-traits". Following initialization with a limited set of seed-traits (two, in most models: see 4.2.2), the population could undergo four types of "events" associated with different probabilities: trait loss and novel trait invention (diametrically opposed); but also trait *modification* and *combination*.

The paper doesn't offer details about how the models would translate into real-world cases, but it seems natural to conceive of modifications as changing *parts*, *details* or *sub-components* of entire traits: one may modify a hammer, for example, by adding a longer handle or changing the inclination or shape of the percussive head.⁵⁴ In any case, if modifications concern parts, sub-components or details of

⁵⁴ Modifications, it seems, may change subordinate parts or aspects of the trait, but by definition they don't alter trait identity (they would be characterized as transformations or novelties otherwise): a hammer stays a hammer despite having a longer handle and differently shaped head. A speculative thought: trait identity seems to correlate more reliably with radical changes in *function*; think of using a hammer as a paperweight.

traits (as it seems correct to presume), we would expect that preserving modifications is conceptualized as *fidelity-I*.

What happens in Laland and Lewis' models is instead rather bizarre (from my framework's perspective, and perhaps also from the perspective of the intuitive fit of the model to reality): modifications within a trait are effectively preserved,⁵⁵ but through originating a new trait unit. Modified or combined traits are registered in the models as *new* variants—i.e. a modified trait counts as an *entire* new trait (Lewis & Laland 2012: 2173). So fidelity-I doesn't feature explicitly, despite talk of modification; it appears somehow to piggyback on, or be unproblematically reabsorbed into, fidelity-II.⁵⁶

What does this mean? Of course, representing modifications as innovations is a legitimate modelling choice and doesn't undermine the model's coherence, but it reflects an assumption about reality which may not be so plausible after all. This underlying assumption is, I think, what I have called the "quantitative view" of accumulation and complexity, which entails a corresponding quantitative conception of trait identity. For Lewis and Laland, a complex or cumulative cultural trait is nothing more than the *additive sum* of its parts: a string of subcomponents with no particular internal organization. It follows that a trait is primarily determined by the quantitative count of its subcomponents, their *number*⁵⁷ (rather than on how they

⁵⁵ Or better: depending on the value assigned to fidelity in the model, they have a certain probability to do so.

⁵⁶ In effect, fidelity is unequivocally conceptualized as TP in this paper. Ultimately, the four possible events in the models (modification, combination, loss, invention) can produce only two outcomes at each iteration: the culture of the group is either "augmented with a new trait, or a lost trait removed" (ibid: 2173). The probability of preservation, directly inverse to that of loss, is the only overarching measure of fidelity.

⁵⁷ More precisely, in quantitative accounts trait identity might depend on the number *and kind* of parts. Suppose that a trait A can undergo modifications, represented by indicized letters ($_{x, y, z}$). A_{xy} is equivalent to A_{yx}; but not to A_{xyz} (variation in number of modifications). But A_{xy} is also *not* equivalent to A_{xr} (no variation in number of modifications, but in kind). In any case, the models are designed so as to keep track of the subsequent addition of modifications in a trait lineage; a lineage with more

are organized or on their interaction); accordingly, a change in the *number* of subcomponents—corresponding to the "modification event" in the model—is interpreted as a *new* trait, rather than as the *same* trait that has undergone a modification.

But what matters more to my argument is the fact that if a trait is nothing more than the sum of its parts, it follows that preserving all of its parts (fidelity-I) will seamlessly translate into preservation of entire trait (fidelity-II), and vice versa (preserving wholes leads automatically to preservation of parts; indeed, there is little reason to distinguish between different levels of fidelity under such a framework, which is exactly what Lewis and Laland do).

However, as I hope to have convincingly argued, this seamless correspondence of types of fidelity is illusory in most real-world settings. Thus, Laland and Lewis' extension of the quantitative perspective to the level of individual cultural traits does not make their account more refined or penetrating, compared to Jablonka and Avital's. Rather, if anything, it opens it to criticism of its applicability.

In conclusion, I think that quantitative interpretations of the RA are bound to be inadequate at whichever level we consider them. Only a quantitative interpretation of L2 is legitimate, but it is trivial: clearly nothing like cumulative culture could exist at all if cultural traits or repertoires systematically failed to be preserved. But as soon as one considers the whole line of reasoning comprising L1 and L2, problems arise. Flattening imitation into an "all-copying" definition essentially silences L1: as per above, this ends up depriving the RA of its explanatory power, reducing it to a descriptive device, and essentially undermining the distinction between culture and cumulative culture.

modifications events is "more complex", confirming the reliance on a quantitative account of complexity.

5.4.1. Other quantitative interpretations: a beam of redemption

I have been quite severe with quantitative interpretations. In this section, I examine a couple of ideas that take a quantitative reading of the RA as their starting point, but introduce a somewhat innovative twist in the argument—specifically, one that was not present, I believe, in Tomasello's original suggestion.

Both these ideas assume a basic meaning of fidelity as type-II: preservation of traits is the crucial fact, and there is not much space for the idea of piecemeal, finegrained transmission of traits that was crucial to gradualist readings (fidelity-I). This is the main reason for classing these ideas together within the quantitative family. In reality, however, they are quite liberally inspired by the RA and do not necessarily presuppose a programmatic conceptualization of complexity or cumulativeness.

The first idea I may call the "patent" interpretation of the RA. In reality, this interpretation is a reworking of L3, the putative RA conclusion: imitation is necessary or particularly conducive for cumulative culture. It then sets out to justify the claim, confining the imitation-fidelity node to a decisively subordinate position.

The crucial element of this idea is that imitation is taken to contrast not so much with *other social learning mechanisms* but with *independent invention*. Copying something that has already worked uses less time and fewer resources than inventing it anew, as the phenomenon of patents vividly illustrates. The spare time and resources can then be devoted to other creative efforts, enlarging the overall pool of cultural traits available. This idea is unrefined, but it has clout. I think it should figure in the list of possible RA interpretations.

The second idea may be called the "creative crowding" interpretation. It explores further the significance of fidelity-II or trait preservation to large, complex cultural repertoires—in short, it focuses on L2 of the RA. Something along these lines may be, in effect, part of Lewis and Laland's overall argument; it may be implicitly present also in Muthukrishna and Henrich's idea of "combinatorial" culture.

The grounding intuition here is that the more traits get preserved, the more they seem to prompt the creation of new traits—perhaps simply by introducing greater probability of recombination, or because they select for superior cultural capacities.

This idea may capture a genuine possibility, but needs substantial chipping away. We need a theory of how and why a crowded cultural scene selects for further crowding or complexity. Certainly, many and more other elements than trait preservation will claim a spot on the stage. So, again, while inspired by the importance of fidelity to cumulative culture, emerging from L2 of the RA, the idea represents a significant extension and reworking of the argument.

5.5. Conclusions

I wish to close this work with a very brief note on its genesis and nature. This doctoral dissertation started out as a broad enquiry of the role of imitation in the evolutionary trajectory of the human lineage. It soon became apparent, however, that "imitation" is a quite liberal name tag for a variety of phenomena. The connection of "imitation" to what I identified in this thesis as the Ratchet Argument, however, appeared virtually ubiquitous in the literature dealing with human evolutionary theories. It was, hence, at first with surprise, and then with growing disconcertion (to say nothing of time expenditure) that I discovered that I was unable to pin and hold down a clear picture of this line of reasoning concerned with imitation.

The chronicle of these clarificatory efforts make up the bulk of this work. While trying to extricate and structure the crucial elements of ratchet hypotheses, I unearthed a variety of fascinating and stimulating intuitions, problems, issues and debates. Many of these have made an incidental and sometimes only brief appearance in this work; I have frequently resisted the urge to investigate details, while endeavouring to keep my navigation bar focussed on an apparently simple but eventually very demanding task: disambiguating the RA. I hope that I have convincingly demonstrated, at the very least, that this argument has been subjected to a variety of competing interpretations (by authors often unaware of the fact that this is what they are doing). Eventually, this work serves as a broad diagnosis of the several ambiguities plaguing the RA formulation, reception and application; the proposed "therapy", unfortunately, was largely and obligatorily left out of the picture.

Nonetheless, some "therapeutic hints" can be quite readily found throughout the text, suggesting which directions are worth pursuing and which are instead less likely to be productive. The RA has been an influential argument, and has enjoyed considerable attention in the nearly thirty years since its appearance, triggering and shaping a lot of empirical and theoretical research. Its underdetermination may have contributed to its fecundity through this time, but the newly-reached maturity of cultural evolution studies now makes an enhanced terminological and conceptual clarity vital.

It is therefore useful to see that what I have called "quantitative interpretations" of the RA are frequently redundant, for they do not add any particular insight to the regular notion of cultural transmission (unless implemented by corollary considerations, like in the "patent" or "creative crowding" readings, which may deserve further examination). It is also important to see that gradualistflavoured versions of the RA may have a point, but also that their applicability is likely to be very limited; this result and its corollaries may be usefully incorporated in the theoretical premises that lead to design cumulative culture experiments (cf. Caldwell and Millen 2008a,b). Ultimately, I think that this work has the merit to expose, if not to resolve, serious gaps in the conceptualization of cultural fidelity. Fidelity has a critical explanatory role in virtually all questions regarding culture (not only cumulative culture): therefore this notion demands serious attention and a thorough restructuring in order to properly serve the future science of cultural evolution.

References

- Acerbi, A., & Mesoudi, A. (2015). If We Are All Cultural Darwinians, what's the fuss about? *Biology and Philosophy* 30 (4). Springer Netherlands: 481–503.
- Acerbi, A., & Tennie, C. (2016). The role of redundant information in cultural transmission and cultural stabilization. *Journal of Comparative Psychology*, *130*(1), 62–70.
- Acerbi, A., Jacquet, P. O., & Tennie, C. (2012). Behavioral constraints and the evolution of faithful social learning. *Current Zoology*, *58*(2), 307–318.
- Acerbi, A., Kendal, J., & Tehrani, J. J. (2017). Cultural complexity and demography: The case of folktales. *Evolution and Human Behavior*, *38*(4), 474–480.
- Acerbi, A., Tennie, C., & Nunn, C. L. (2011). Modeling imitation and emulation in constrained search spaces. *Learning & Behavior*, *39*(2), 104–114.
- Aisner, R., & Terkel, J. (1992). Ontogeny of pine cone opening behaviour in the black rat, Rattus rattus. *Animal Behaviour*, 44, Part 2(0), 327–336.
- Akins, C. K., & Zentall, T. R. (1996). Imitative learning in male Japanese quail (Coturnix japonica) using the two-action method. *Journal of Comparative Psychology (Washington, D.C. : 1983), 110*(3), 316–320.
- Alberch, P. (1991). From Genes to Phenotype: Dynamical Systems and Evolvability. *Genetica* 84 (1). Springer: 5–11.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). DNA replication, repair, and recombination. *Molecular Biology of the Cell*, 235-97.
- Albertson, T. M., & Preston, B. D. (2006). DNA Replication Fidelity: Proofreading in Trans. *Current Biology*, *16*(6), R209–R211.
- Amundson, R., & Lauder, G. V. (1994). Function without purpose. *Biology & Philosophy*, *9*(4), 443–469.

Andersson, C. (2013). Fidelity and the emergence of stable and cumulative

sociotechnical systems. PalaeoAnthropology, 2013, 88–103.

- Andersson, C., Read, D., Eriksson, K., Gilligan, I., O'Brien, M. J., Alexander
 Bentley, R., Taylor, T., Vaesen, K. (2016). The evolution of cultural
 complexity: not by the treadmill alone. *Current Anthropology*, *57*(3), 261–286.
- Anon. (2005). So much more to know. Science, (309), 78–102.
- Arbib, M. A. (2011). From mirror neurons to complex imitation in the evolution of language and tool use. *Annual Review of Anthropology*, *40*, 257–273.
- Ariew, A., Cummins, R., & Perlman, M. (Eds.) (2002) *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford University Press, USA.
- Atran, S. (2001). The trouble with memes Inference versus Imitation in Cultural Creation. *Human Nature*, *12*(4), 351–381.
- Atran, S. (2002). In gods we trust: The evolutionary landscape of religion. Oxford University Press.
- Atran, Scott. 2001. "The Trouble with Memes." Human Nature 12 (4): 351–81.
- Averbeck, B. B., & Lee, D. (2004). Coding and transmission of information by neural ensembles. *Trends in Neurosciences*, 27(4), 225–230.
- Avital, E., & Jablonka, E. (2000). *Animal Traditions: Behavioral Inheritance in Evolution*. Cambridge, UK: Cambridge University Press.
- Baktir, H. (2003). The Concept of Imitation in Plato and Aristotle (Aristo ve Plato'da Taklit). *Sosyal Bilimler Enstitusu Dersiqi Sayi*, *2*(2), 167–179.
- Balfour, H. (1906). Preface. In A. H. L.-F. Pitt-Rivers (Ed.), *The evolution of culture: and other essays*. Clarendon Press.
- Basalla, G. (1988). The evolution of technology. Cambridge University Press.
- Bates, L. A., & Byrne, R. W. (2010). Imitation: What animal imitation tells us about animal cognition. Wiley Interdisciplinary Reviews: Cognitive Science, 1(5)
- Bechtel, W. (2008a). Mechanisms in cognitive psychology: what are the operations? *Philosophy of Science*, *75*(5), 983–994.
- Bechtel, W. (2008b). Mental Mechanisms: Philosophical Perspectives on

Cognitive Neuroscience. New York: Routledge.

- Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36(2), 421–441.
- Beckner, M. (1969). Function and teleology. *Journal of the History of Biology*, 2(1), 151–164.
- Belfiore, E. (1984). A theory of imitation in Plato's Republic. *Transactions of the American Philological Association (1974-), 114, 121–146.*
- Benitez-Bribiesca, L. (2001). Memetics: A dangerous idea. *Interciencia*, 26(1), 29– 31.
- Bergstrom, C. T., & Rosvall, M. (2011). The Transmission Sense of Information.Biology & Philosophy 26 (2). Springer Netherlands: 159–76.
- Bernstein, N. (1967). The Co-Ordination and Regulation of Movements. *The Co-Ordination and Regulation of Movements*. Pergamon Press.
- Binford, L. R. (1980). Willow smoke and dogs' tails: hunter-gatherer settlement systems and archaeological site formation. *American Antiquity*, 45(1), 4–20.

Blackmore, S. (1999). The meme machine (Vol. 25). Oxford University Press.

- Boesch, C. (2003). Is culture a golden barrier between human and chimpanzee? Evolutionary Anthropology: Issues, News, and Reviews, 12(2), 82–91.
- Boesch, C., & Boesch, H. (1990). Tool use and tool making in wild chimpanzees. Folia Primatologica, 54(1–2), 86–99.
- Bourrat, P. (2014a). From Survivors to Replicators: Evolution by Natural Selection Revisited. *Biology & Philosophy* 29 (4): 517–38.
- Bourrat, P. (2014b). How to Read 'heritability'in the Recipe Approach to Natural Selection. *The British Journal for the Philosophy of Science*, *66*(4), 883-903.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago, IL: University of Chicago Press.
- Boyd, R., & Richerson, P. J. (1996). Why culture is common, but cultural evolution is rare. In *Proceedings-British Academy* (Vol. 88, pp. 77–94). Oxford University Press.
- Boyd, R., Richerson, P. J., & Henrich, J. (2013). The cultural evolution of technology: facts and theories. *Cultural Evolution: Society, Technology, Language, and Religion*, 119–142.
- Boyer, P. (1994). *The naturalness of religious ideas: A cognitive theory of religion*. University of California Press.
- Boyer, P. (1998). Cognitive tracks of cultural inheritance: How evolved intuitive ontology governs cultural transmission. *American Anthropologist*, *100*(4), 876–889.
- Brass, M., & Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences*, *9*(10), 489–495. h
- Brown, R. L. (2013). Learning, evolvability and exploratory behaviour: Extending the evolutionary reach of learning. *Biology and Philosophy*, *28*(6), 933–955.
- Brown, R. L. (2014). Identifying behavioral novelty. *Biological Theory*, *9*(2), 135–148.
- Brown, R. L. (2017). Animal traditions. What they are, and why they matter. *The Routledge Handbook of Philosophy of Animal Minds*. Taylor & Francis, 362.

Bugnyar, T., & Huber, L. (1997). Push or pull: an experimental study on imitation

Burrows, N. (1986). Backburning in forest areas. Australian Department of Conservation & Land Management. Retrieved at:

https://science.dpaw.wa.gov.au/people/?sid=9

- Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2007). Enculturated chimpanzees imitate rationally. *Developmental Science*, *10*(4), 31–38.
- Butterfill, S. A., & Sinigaglia, C. (2014). Intention and Motor Representation in Purposive Action. *Philosophy and Phenomenological Research*, 88(1), 119– 145.
- Byrne, R. W. (1999). Imitation without intentionality. Using string parsing to copy the organization of behaviour. *Animal Cognition*, *2*(2), 63–72.
- Byrne, R. W. (2002). Imitation of novel complex actions: what does the evidence from animals mean? *Advances in the Study of Behavior*, *31*, 77–104.
- Byrne, R. W. (2003). Imitation as behaviour parsing. Philosophical Transactions

of the Royal Society of London. Series B: Biological Sciences, 358(1431), 529–536.

- Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: a hierarchical approach. *Behavioral and Brain Sciences*, *21*(21), 667-684-721.
- Byrne, R. W., Tomasello, M., (1995). Do rats ape? *Animal Behaviour*, *50*(5), 1417–1420.
- Caggiano, V., Fogassi, L., Rizzolatti, G., Casile, A., Giese, M. A., & Thier, P. (2012). Mirror neurons encode the subjective value of an observed action. *Proceedings of the National Academy of Sciences*, *109*(29), 11848–11853.
- Calcott, B., & Sterelny, K. (Eds.) (2011). *The Major Transitions in Evolution Revisited*. MIT Press.
- Calcott, B. (2009). Lineage Explanations: Explaining How Biological Mechanisms Change. *The British Journal for the Philosophy of Science* 60 (1): 51–78.
- Caldwell, C. A., & Millen, A. E. (2008a). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior*, *29*(3), 165–171.
- Caldwell, C. A., & Millen, A. E. (2008b). Review. Studying cumulative cultural evolution in the laboratory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1509), 3529–39.
- Caldwell, C. A., & Millen, A. E. (2009). Social Learning Mechanisms and Cumulative Cultural Evolution Is Imitation Necessary? *Psychological Science*, 20(12), 1478–1483.
- Caldwell, C. A., Atkinson, M., & Renner, E. (2016). Experimental Approaches to Studying Cumulative Cultural Evolution. *Current Directions in Psychological Science*, *25*(3), 191–195.
- Call, J., & Carpenter, M. (2002). Three Sources of Information in Social Learning.In K. Dautenhahn & C. L. Nehaniv (Eds.), *Imitation in Animals and Artifacts*.Cambridge, MA: MIT Press.
- Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: a quantitative approach*. (No.16). Princeton University Press.

- Charbonneau, M. (2014). Populations without reproduction. *Philosophy of Science*, *81*(5), 727–740.
- Charbonneau, M. (2015). All innovations are equal, but some more than others: (Re)integrating modification processes to the origins of Cumulative Culture. *Biological Theory*, 10(4), 322–335.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception– behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893.
- Chartrand, T. L., & Van Baaren, R. (2009). Human mimicry. Advances in Experimental Social Psychology, 41, 219–274.
- Claidière, N., & Sperber, D. (2007). The Role of Attraction in Cultural Evolution. Journal of Cognition and Culture 7 (1): 89–111.
- Claidière, N., & Sperber, D. (2009). Imitation explains the propagation, not the stability of animal culture. *Proceedings of the Royal Society B: Biological Sciences*, 277(1681), rspb20091615.
- Claidière, N., & Sperber, D. (2010). The natural selection of fidelity in social learning. *Communicative & Integrative Biology*, *3*(4), 350–351.
- Claidière, N., Scott-Phillips, T.C., & Sperber, D. (2014) How Darwinian Is Cultural Evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1642): 20130368.
- Clay, Z., & Tennie, C. (2017). Is Overimitation a Uniquely Human Phenomenon? Insights From Human Children as Compared to Bonobos. *Child Development*, (July).
- Cole, M., & Engeström, Y. (1993). A cultural-historical approach to distributed cognition. *Distributed Cognitions: Psychological and Educational Considerations*, 1–46.
- Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: from origin to function. *Behavioral and Brain Sciences*, *37*(2), 177–192.
- Cook, R., Bird, G., Lünser, G., Huck, S., & Heyes, C. (2012). Automatic imitation in a strategic context: players of rock-paper-scissors imitate opponents'

gestures. *Proceedings of the Royal Society B: Biological Sciences, 279*(1729), 780–786.

- Daboo, J. (2010). Ritual, rapture and remorse: a study of tarantism and pizzica in Salento. Peter Lang.
- Dawkins, R. (1976). The Selfish Gene. Oxford: Oxford University Press.
- Dawkins, R. (1983). Universal darwinism. In D. S. Bendall (Ed.), *Evolution from Molecules to Man*. Cambridge University Press, New York. Ma.

Dawkins, R. (1986). The Blind Watchmaker. New York: Norton.

- Dean, G. L., Kendal, R. L., Schapiro, S. J., Thierry, B., Laland, K. N., L. G. Dean, R. L. Kendal, S. J. Schapiro, B. Thierry, K. N. L., Laland, K. N. (2012). Identification of the social and cognitive processes underlying human cumulative culture. *Science*, 335(March), 1114–1117.
- Dean, L. G., Vale, G. L., Laland, K. N., Flynn, E., & Kendal, R. L. (2014). Human cumulative culture: A comparative perspective. *Biological Reviews*, *89*(2), 284–301.
- Dennett, D. C. (1993). Consciousness Explained. London: Penguin UK.
- Dennett, D. C. (1995). Darwin's dangerous idea. Simon & Schuster.
- Dennett, D. C. (1997). Appraising Grace: what evolutionary good is God? (Review of Walter Burkert, Creation of the Sacred: Tracks of Biology in Early Religions). *The Sciences*, 37(1), 39–44.
- Dijksterhuis, A. (2005). Why we are social animals: The high road to imitation as social glue. *Perspectives on Imitation: From Neuroscience to Social Science*, 2, 207–220.
- Drake, J. W., Charlesworth, B., Charlesworth, D., & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics*, *148*(4), 1667–1686.
- Else, G. F. (1958). "Imitation" in the Fifth Century. *Classical Philology*, *53*(2), 73–90.
- Enquist, M., Ghirlanda, S., Jarrick, A., & Wachtmeister, C.-A. (2008). Why does human culture increase exponentially? *Theoretical Population Biology*, *74*(1), 46–55.

- Enquist, M., Strimling, P., Eriksson, K., Laland, K., & Sjostrand, J. (2010). One cultural parent makes no culture. *Animal Behaviour*, *79*(6), 1353–1362.
- Enquist, M., Ghirlanda, S. & Eriksson, K. (2011). Modelling the Evolution and Diversity of Cumulative Culture. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366 (1563): 412–23.
- Fawcett, T. W., Skinner, A. M. J., & Goldsmith, A. R. (2002). A test of imitative learning in starlings using a two-action method with an enhanced ghost control. *Animal Behaviour*, 64(1965), 547–556.
- Fernandez-Leiro, R., Conrad, J., Yang, J.-C., Freund, S. M. V, Scheres, S. H. W., & Lamers, M. H. (2017). Self-correcting mismatches during high-fidelity DNA replication. *Nature Structural & Molecular Biology*.
- Fiorito, G., & Scotto, P. (1992). Observational learning in Octopus vulgaris. *Science*, *256*(5056), 545–547.
- Fragaszy, D. M., & Perry, S. (2008). *The biology of traditions: models and evidence*. Cambridge University Press.
- French, L., Sutherland, R. & Garry, M. (2006) Discussion Affects Memory for True and False Childhood Events. *Applied Cognitive Psychology* 20 (5). Wiley Online Library: 671–80.
- Frenken, K. (2006). Technological innovation and complexity theory. *Economics* of Innovation & New Technology, 15(2), 137–155.
- Friedman, M. (1974). Explanation and scientific understanding. *The Journal of Philosophy*, *71*(1), 5–19.
- Galef, B. G. (1988). Imitation in animals: history, definition, and interpretation of data from the psychological laboratory. In *Social learning: Psychological and biological perspectives* (Vol. 28, pp. 3–28). Hillsdale, New Jersey: Erlbaum.
- Galef, B. G. (1990). Tradition in animals: Field observations and laboratory analyses. In Beckoff & Jamieson (Eds.), *Interpretation and explanation in the study of animal behavior* (pp. 74–95). Boulder: Westview Press.
- Galef, B. G. (1992). The question of animal culture. *Human Nature*, *3*(2), 157–178.

- Galef, B. G. (1998). Recent progress in studies of imitation and social learning in animals. *Advances in Psychological Science*, *2*, 275–300.
- Galef, B. G. (2012). Social learning and traditions in animals: evidence,
 definitions, and relationship to human culture. Wiley Interdisciplinary
 Reviews: Cognitive Science, 3(6), 581–592.
- Galef, B. G. (2013). Imitation and local enhancement: Detrimental effects of consensus definitions on analyses of social learning in animals. *Behavioural Processes*, 100, 123–130.
- Galef, B. G. (2014). Social Learning, tradition, and culture: data and debate. In K. Yasukawa & Z. Tang-Martinez (Eds.), *Animal Behaviour How and Why Animals Do the Things They Do* (pp. 115–150). Santa Barbara: ABC-CLIO.
- Galef, B. G. (2015). Laboratory studies of imitation / field studies of tradition.
 Towards a synthesis in animal social learning. *Behavioural Processes*, *112*, 114–119.
- Gallese, V., Gernsbacher, M. a., Heyes, C., Hickok, G., & Iacoboni, M. (2011). Mirror Neuron Forum. *Perspectives on Psychological Science*, *6*(4), 369–407.
- Gergely, G. G., & Csibra, G. (2006). Sylvia' s recipe : The role of imitation and pedagogy in the transmission of cultural knowledge. *Roots of Human Sociality: Culture, Cognition, and Human Interaction,* 229–255.
- Gerrans, P. (2013). Imitation, mind reading, and social learning. *Biological Theory*, *8*(1), 20–27.
- Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, *69*(S3), S342–S353.
- Glennan, S. (2017). *The new mechanical philosophy*. Oxford: Oxford University Press.
- Glock, H. (2012). The anthropological difference: what can philosophers do to identify the differences between human and non-human animals? *Royal Institute of Philosophy Supplement*, *70*, 105–131.
- Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. New York: Oxford University Press.

- Griesemer, J., Haber, M. H., Yamashita, G., & Gannett, L. (2005). Critical Notice: Cycles of contingency–developmental systems and evolution. *Biology and Philosophy*, *20*(2), 517–544.
- Griesemer, J.R. 2014. "Reproduction and the Scaffolded Development of Hybrids." *Developing Scaffolds in Evolution, Culture, and Cognition*, 23–55.
- Griesemer, J. R. (2005). The Informational Gene and the Substantial Body: On the Generalization of Evolutionary Theory by Abstraction. *Idealization XII: Correcting the Model. Idealization and Abstraction in the Sciences* 86: 59– 115.
- Griesemer, J. (2000). Development, Culture, and the Units of Inheritance. *Philosophy of Science* 67: S348–68.
- Griffiths, P., & Gray, R. (2001). Darwinism and developmental systems. In
 Oyama, S.Griffiths, P. & Gray, R. (Eds.), *Cycles of contingency: developmental systems and evolution* (pp. 195–218). Cambridge, Mass.: MIT
 Press.
- Haidle, M. N., & Conard, N. J. (2011). The nature of culture. *Mitteilungen Der Gesellschaft Für Urgeschichte*, 20, 65.
- Henrich, J. (2004). Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses, the Tasmanian case. *American Antiquity*, *69*(2), 197–214.
- Henrich, J. (2015). The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter. Princeton University Press.
- Henrich, J., & Boyd, R. (2002). On modeling cognition and culture. *Journal of Cognition and Culture*, *2*(2), 87–112.
- Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. Evolutionary Anthropology: Issues, News, and Reviews, 12(3), 123–135.
- Henrich, J., Boyd, R., & Richerson, P. J. (2008). Five misunderstandings about cultural evolution. *Human Nature*, *19*(2), 119–137.
- Henrich, J., Boyd, R., Derex, M., Kline, M. A., Mesoudi, A., Muthukrishna, M.,

Powell, A., Shennan, J., & Thomas, M. G. (2016). Understanding cumulative cultural evolution. *Proceedings of the National Academy of Sciences*, *113*(44), E6724-E6725.

Heyes, C. (1993). Imitation, culture and cognition. *Animal Behaviour*. 46(5), 999-1010.

- Heyes, C. (1994a). Imitation and Culture: Longevity, Fecundity and Fidelity in Social Transmission. *Behavioral Aspects of Feeding*.
- Heyes, C. (1994b). Social learning in animals: categories and mechanisms. Biological Reviews, 69(2), 207–231.
- Heyes, C. (1995). Imitation and flattery: a reply to Byrne & Tomasello. *Animal Behaviour*, *50*(5), 1421–1424.
- Heyes, C. (1996). Introduction: identifying and defining imitation. In *Social Learning in Animals: The Roots of Culture* (pp. 211–220). Academic Press.

Heyes, C. (2010a). Mesmerising mirror neurons. NeuroImage, 51(2), 789–791.

- Heyes, C. (2010b). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, *34*(4), 575–583.
- Heyes, C. (2012a). What's social about social learning? *Journal of Comparative Psychology*, *126* (2)(May 2012), 193–202.

Heyes, C. (2012b). Grist and mills: on the cultural origins of cultural learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1599), 2181–2191.

- Heyes, C. (2013a). A new approach to mirror neurons: developmental history, system-level theory and intervention experiments. *Cortex*, *49*(10), 2946–2948.
- Heyes, C. (2013b). What can imitation do for cooperation? In K. Sterelny, R.
 Joyce, B. Calcott, & B. Fraser (Eds.), *Cooperation and Its Evolution* (pp. 314–342). Cambridge, MA: The MIT Press.
- Heyes, C. (2017). When does social learning become cultural learning? Developmental Science, 20(2).

Heyes, C. (2016). Imitation: Not in Our Genes. Current Biology, 26(10), R412-

R414.

- Heyes, C. (2018). *Cognitive gadgets: the cultural evolution of thinking*. Harvard University Press.
- Heyes, C., & Saggerson, A. (2002). Testing for imitative and nonimitative social learning in the budgerigar using a two-object/two-action test. *Animal Behaviour*, 64(6), 851–859.
- Heyes, C., Dawson, G. R., & Nokes, T. (1992). Imitation in rats: Initial responding and transfer evidence. *The Quarterly Journal of Experimental Psychology Section B*, 45(3), 229–240.
- Hickok, G., & Hauser, M. (2010). (Mis) understanding mirror neurons. *Current Biology*, *201*, 613–618.
- Hinde, R. A., & Fisher, J. (1951). Further observations on the opening of milk bottles by birds. *British Birds*, *44*(12), 393–396.
- Hiscock, P. (2014). Learning in Lithic Landscapes: A Reconsideration of the Hominid "Toolmaking" Niche. *Biological Theory*, *9*(1), 27–41.
- Hoppitt, W., & Laland, K. N. (2013). *Social learning: an introduction to mechanisms, methods, and models*. Princeton University Press.
- Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens).
 Animal Cognition, 8(3), 164–181.
- Horner, V., Whiten, A., Flynn, E. G., & de Waal, F. B. M. (2006). Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proceedings of the National Academy of Sciences* of the United States of America, 103(37), 13878–83.
- Huang, C. T., & Charman, T. (2005). Gradations of emulation learning in infants' imitation of actions on objects. *Journal of Experimental Child Psychology*, 92(3), 276–302.
- Hunt, G. R. (2014). New Caledonian crows' (Corvus moneduloides) Pandanus tool designs: Diversification or Independent Invention? *The Wilson Journal of Ornithology*, 126(1), 133–139.

- Hunt, G. R., & Gray, R. D. (2003). Diversification and cumulative evolution in New Caledonian crow tool manufacture. *The Royal Society*, *270*(1517), 867–874.
- Hurley, S., & Chater, N. (Eds.) (2005). *Perspective on Imitation: From Neuroscience to Social Science.* (2 Vols.). Cambridge, MA: MIT press.
- Iacoboni, M. (2009). Imitation, Empathy, and Mirror Neurons. *Annual Review of Psychology*, *60*(1), 653–670.
- Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, *2*(1), 119–135.
- Jablonka, E. (2002a). Animal Traditions: Behavioral Inheritance in Evolution. *Trends in Ecology & Evolution*, *17*(4), 198.
- Jablonka, E. (2002b). Information: Its Interpretation, Its Inheritance, and Its Sharing. *Philosophy of Science*, *69*(4), 578–605.
- Jablonka, E., & Lamb, M. (2005). Evolution in Four Dimension: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life. Cambridge, MA: MIT Press.
- Jablonka, E., & Lamb, M. (2006). The evolution of information in the major transitions. *Journal of Theoretical Biology*, *239*(2), 236–246.
- Johnson, A., & Proctor, R. W. (2004). Attention: Theory and practice. Sage.
- Jones, S. S. (2007). Imitation in infancy: The development of mimicry. *Psychological Science*, *18*(7), 593–599.
- Kauffman, S., & Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, *128*(1), 11–45.
- Kauffman, S., & Macready, W. (1995). Technological Evolution and Adaptive Organizations. *Complexity*, 1(2), 26–43.
- Kawai, M. (1965). Newly-acquired pre-cultural behavior of the natural troop of Japanese monkeys on Koshima Islet. *Primates*, *6*(1), 1–30.
- Kawakami, K., Dovidio, J. F., & Dijksterhuis, A. (2003). Effect of social category priming on personal attitudes. *Psychological Science*, *14*(4), 315–319.
- Kawamura, S. (1959). The process of sub-culture propagation among Japanese

macaques. *Primates*, 2(1), 43-60.

- Kis, A., Huber, L., & Wilkinson, A. (2014). Social learning by imitation in a reptile (Pogona vitticeps). *Animal Cognition*, 18(1), 325–331.
- Kobayashi, Y., Ohtsuki, H., & Wakano, J. Y. (2016). Population size vs. social connectedness—a gene-culture coevolutionary approach to cumulative cultural evolution. *Theoretical Population Biology*, *111*, 87–95.
- Kohler, E. (2002). Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science*, *297*(5582), 846–848.
- Kroeber, A. L., & Kluckhohn, C. (1952). Culture: A critical review of concepts and definitions. Papers. Peabody Museum of Archaeology & Ethnology, Harvard University.
- Kuhn, S. L. (1992). On planning and curated technologies in the Middle Paleolithic. *Journal of Anthropological Research*, *48*(3), 185–214.
- Kunkel, T. A. (2004). DNA replication fidelity. *Journal of Biological Chemistry*, 279(17), 16895–16898.
- Kunkel, T. A., & Bebenek, K. (2000). DNA Replication Fidelity. Annual Review of Biochemistry, 69(1), 497–529.
- Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, *14*(4), 334–339.
- Laland, K. N., & Galef, B. G. (2009). *The Question of Animal Culture*. Harvard University Press.
- Laland, K. N., & Hoppitt, W. (2003). Do animals have culture? *Evolutionary Anthropology: Issues, News, and Reviews, 12*(3), 150–159.
- Laland, K. N., Kendal, J. R., & Kendal, R. L. (2009). Animal culture: problems and solutions. *The Question of Animal Culture*, 174–197.
- Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., Odling-Smee, J. (2015). The extended evolutionary synthesis: its structure, assumptions and predictions. In *Proceedings of the Royal Society Bulletin* (Vol. 282, p. 20151019).

Lefebvre, L. (2013). Brains, innovations, tools and cultural transmission in birds,

non-human primates, and fossil hominins. *Frontiers in Human Neuroscience*, 7(June), 245.

- Lefebvre, L., & Bouchard, J. (2003). Social learning about food in birds. *The Biology of Traditions: Models and Evidence*, 94–126.
- Lepage, J., & Théoret, H. (2007). The mirror neuron system: grasping others' actions from birth? *Developmental Science*, *10*(5), 513–523.
- Levy, A. (2011). Information in biology: A fictionalist account. *Nous, 45*(4), 640–657.
- Lewens, T. (2002). Darwinnovation! *Studies in History and Philosophy of Science*, 33(1), 199–207.
- Lewis, H. M., & Laland, K. N. (2012). Transmission fidelity is the key to the buildup of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2171–2180.
- Lewontin, R. (1985). Adaptation. In R. Levins & R. Lewontin (Eds.), *The Dialectical Biologist* (pp. 65–84). Cambridge, MA: Harvard University Press.
- Loeb, L. A., & Kunkel, T. A. (1982). Fidelity of DNA synthesis. *Annual Review of Biochemistry*, *51*(1), 429–457.
- Lombard, M. (2016). Mountaineering or ratcheting? Stone Age hunting weapons as proxy for the evolution of human technological, behavioral and cognitive flexibility. In *The nature of culture* (pp. 135–146). Springer.
- Lumsden, C. J., & Wilson, E. O. (1981). *Genes, minds and culture*. Cambridge, MA: Harvard University Press.
- Lycett, S. J., & Norton, C. J. (2010). A demographic model for Palaeolithic technological evolution: The case of East Asia and the Movius Line. *Quaternary International*, 211(1-2), 55–65.
- Lycett, S. J., Schillinger, K., Kempe, M., & Mesoudi, A. (2015). Learning in the Acheulean: Experimental Insights Using Handaxe Form as a "Model Organism." In *Learning strategies and cultural evolution during the Palaeolithic* (pp. 155–166). Springer.

Lyons, D. E., Damrosch, D. H., Lin, J. K., Macris, D. M., & Keil, F. C. (2011). The

scope and limits of overimitation in the transmission of artefact culture. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 366*(1567), 1158–1167.

- Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences*, *104*(50), 19751–19756.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1–25.
- Mameli, M. (2004). Nongenetic Selection and Nongenetic Inheritance, British Journal for the Philosophy of Science 55 (1): 35–71.
- Mayr, E. (1998). The Multiple Meanings of "Teleological." *History and Philosophy* of the Life Sciences, 20(1), 35–40.
- Mayr, E. (1994). Typological versus population thinking. *Conceptual Issues in Evolutionary Biology*, 157–160.
- McCormack, T., Hoerl, C., Butterfill, S., & Introduction, A. (2011). Tool use and causal cognition, *123*(October), 1–17.
- McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to supercopying: Adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *British Journal of Psychology*, *102*(1), 1–18.
- Meltzoff, A. N. (1988). The human infant as Homo imitans. In T. R. Zentall & B. G. Galef (Eds.), Social learnig: Psychological and biological perspectives (p. pp.319-341). Hillsdale, NJ: Elrbaum.
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*(4312), 75–78.
- Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, 702–709.
- Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development & Parenting*, *6*(3–4), 179.
- Meselson, M., & Stahl, F. W. (1958). The replication of DNA in Escherichia coli. Proceedings of the National Academy of Sciences, 44(7), 671–682.

Mesoudi, A. (2009). How cultural evolutionary theory can inform social psychology and vice versa. *Psychological Review*, *116*(4), 929.

- Mithen, S. (1999). Imitation and cultural change: a view from the Stone Age, with specific reference to the manufacture of handaxes. *Mammalian Social Learning: Comparative and Ecological Perspectives*, (72), 389.
- Morikawa, K., Agarwal, S., Elkan, C., & Cottrell, G. (2001). A taxonomy of computational and social learning. In *Proc. of Workshop on Developmental Embodied Cognition* (Vol. 557).

Morin, O. (2015). How Traditions Live and Die. Oxford University Press.

- Morin, O. (2013). How Portraits Turned Their Eyes upon Us: Visual Preferences and Demographic Change in Cultural Evolution. *Evolution and Human Behavior* 34 (3). Elsevier: 222–29.
- Muthukrishna, M., & Henrich, J. (2016). Innovation in the collective brain. *Phil. Trans. R. Soc. B*, *371*(1690), 20150192.
- Nehaniv, C. L., & Dautenhahn, K. (2001). Like me?-measures of correspondence and imitation. *Cybernetics & Systems*, *32*(1–2), 11–51.
- Nehaniv, C. L., & Dautenhahn, K. (2007). Imitation and Social Learning in Robots, Humans and Animals. (C. L. Nehaniv & K. Dautenhahn, Eds.). Cambridge, MA: Cambridge University Press.
- Nguyen, N. H., Klein, E. D., & Zentall, T. R. (2005). Imitation of a two-action sequence by pigeons. *Psychonomic Bulletin & Review*, *12*(3), 514–518.
- Nielsen, M. (2006). Copying actions and copying outcomes: social learning through the second year. *Developmental Psychology*, 42(3), 555–65.
- Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari Bushman Children and the Origins of Human Cultural Cognition. *Psychological Science*, 21(5), 729–736.
- Nielsen, M., Subiaul, F., Galef, B., Zentall, T., & Whiten, A. (2012). Social learning in humans and nonhuman animals: theoretical and empirical dissections. *Journal of Comparative Psychology (Washington, D.C. : 1983)*, 126(2), 109– 13.

- Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari Bushman Children and the Origins of Human Cultural Cognition. *Psychological Science* 21 (5): 729–36.
- Norman, D. (2013). The design of everyday things: Revised and expanded edition. Basic Books (AZ).
- Oakley, K. P. (1957). Man the Tool-maker. Chicago: University of Chicago Press.
- Odling-smee, J., Kendal, J.R., & Tehrani. J.J. (2011). Human Niche Construction, *366*(1566): 783–934.
- Odling-Smee, J. (2007). Niche Inheritance: A Possible Basis for Classifying Multiple Inheritance Systems in Evolution. *Biological Theory* 2 (3): 276–89.
- Oswalt, W. H. (1972). *Habitat and technology: the evolution of hunting*. Holt, Rinehart and Winston.

Oswalt, W. H. (1976). An anthropological analysis of food-getting technology. Wiley.

- Packer, C., & Ruttan, L. (1988). The evolution of cooperative hunting. *The American Naturalist*, 132(2), 159–198.
- Pavlov, Y. I., Frahm, C., McElhinny, S. A. N., Niimi, A., Suzuki, M., & Kunkel, T. A. (2006). Evidence that errors made by DNA polymerase α are corrected by DNA polymerase δ. *Current Biology*, *16*(2), 202–207.

Pepperberg, I. (2005). Insights into Vocal Imitation in African Grey Parrot (Psitthacus erithacus). In S. Hurley & N. Chater (Eds.), Perspectives on Imitation: From Neuroscience to Social Science. Vol. 1: Mechanism of Imitation and Imitation in Animals (Vol. I). Cambridge, MA: MIT press.

- Perry, S. (2011). Social traditions and social learning in capuchin monkeys (Cebus). Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1567), 988–996.
- Pradhan, G. R., Tennie, C., & van Schaik, C. P. (2012). Social organization and the evolution of cumulative technology in apes and hominins. *Journal of Human Evolution*, 63(1), 180–190.

Querbes, A., Vaesen, K., & Houkes, W. (2014). Complexity and demographic

explanations of cumulative culture. PLoS ONE, 9(7).

- Ramachandran, V. S. (2000). Mirror neurons and imitation learning as the driving force behind "the great leap forward" in human evolution. Edge New York, NY.
- Ramsey, G. (2013). Culture in humans and other animals. *Biology & Philosophy*, 28(3), 457–479.
- Ramsey, G. (2017). What is animal culture? In K. Andrews & J. Beck (Eds.), *The Routledge handbook of philosophy of animal minds*. Taylor & Francis.
- Ray, E., & Heyes, C. (2011). Imitation in infancy: the wealth of the stimulus. Developmental Science, 14(1), 92–105.
- Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W.,
 Fogarty, L., Ghirlanda, S., Lillicrap, T. & Laland, K. N. (2010). Why Copy
 Others? Insights from the Social Learning Strategies Tournament. *Science* 328 (2010): 208–13.
- Richerson, P. J., & Boyd, R. (2005). *Not by Genes Alone*. Chicago: The University of Chicago Press.
- Ridley, M. (2000). *Mendel's Demon. Gene justice and the complexity of life.* London: Weidenfeld & Nicholson.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2(September), 661–70.

Ruben, D. (1993). Explanation. Paris: Oxford University Press.

- Saleh, N., Scott, A. G., Bryning, G. P., & Chittka, L. (2007). Distinguishing signals and cues: bumblebees use general footprints to generate adaptive behaviour at flowers and nest. *Arthropod-Plant Interactions*, 1(2), 119–127.
- Seel, N. M. (2012). Preface. In *Encyclopedia of the Sciences of Learning* (1st ed.). Springer US.
- Seppänen, J.-T., Forsman, J. T., Mönkkönen, M., Krams, I., & Salmi, T. (2011). New behavioural trait adopted or rejected by observing heterospecific tutor fitness. *Proc. R. Soc. B*, 278(November 2010), 1736–1741.

Shea, N. (2009). Imitation as an inheritance system. *Philosophical Transactions* of the Royal Society B: Biological Sciences, 364(1528), 2429–2443.

Sherry, D. F., & Galef Jr, B. G. (1984). Cultural transmission without imitation: milk bottle opening by birds. *Animal Behaviour*, *32*(3), 937–938.

Sherry, D. F., & Galef Jr, B. G. (1990). Social learning without imitation: more about milk bottle opening by birds. *Animal Behaviour*, 40(5), 987–989.

Shipton, C. (2010). Imitation and Shared Intentionality in the Acheulean. *Cambridge Archaeological Journal*, 20(2), 197–210.

Shipton, C., Petraglia, M., & Paddayya, K. (2009). Inferring Aspects of Acheulean Sociality and Cognition from Lithic Technology. In *Lithic Materials and Paleolithic Societies* (pp. 217–231). Oxford: Wiley-Blackwell.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*, 467–482.

Simon, M., & Goes, J. (2013). Assumption, limitations, delimitations, and scope of the study. In *Dissertation and scholarly research: Recipes for success* (pp. 1–3). Seattle: Dissertation Success, LLC.

Smaldino, P. E. (2013). The cultural evolution of emergent group-level traits. Behavioral and Brain Sciences.

Solée, R. V, Valverde, S., Casals, M. R., Kauffman, S. A., Farmer, D., & Eldredge,
 N. (2013). The evolutionary ecology of technological innovations.
 Complexity, 18(4), 15–27.

Spence, K. W. (1937). Experimental studies of learning and the higher mental processes in infra-human primates. *Psychological Bulletin*, *34*(10), 806.

Sperber, D. (1996). *Explaining culture: a naturalistic approach*. Oxford, UK: Blackwell.

Sterelny, K. (2001a). Explaining Culture: A Naturalistic Approach. Dan Sperber. Mind 110 (439): 845–54.

Sterelny, K. (2001b). Niche construction, developmental systems, and the extended replicator. In Cycles of Contingency: Developmental Systems and Evolution. MIT Press, Cambridge, MA (pp. 333–350). Cambridge, Mass.: MIT Press.

- Sterelny, K. (2003). *Thought in a hostile world*. Malden, MA: Blackwell Publishing.
- Sterelny, K. (2004). Symbiosis, Evolvability and Modularity. In G. Schlosser & G.
 P. Wagner (Eds.), *Modularity in development and evolution*. University of Chicago Press.
- Sterelny, K. (2006). The evolution and evolvability of culture. *Mind and Language*, *21*(2), 137–165.
- Sterelny, K. (2007). SNAFUS : An Evolutionary Perspective, 2(3), 317–328.
- Sterelny, K. (2009). Peacekeeping in the culture wars. *The Question of Animal Culture*, 288–304.
- Sterelny, K. (2011). Evolvability Reconsidered. In B. Calcott & K. Sterelny (Eds.), *The Major Transitions in Evolution Revisited* (Vienna ser). MIT Press.
- Sterelny, K. (2012). The Evolved Apprentice. Cambridge, MA: MIT Press.
- Sterelny, K. (2016). Cooperation, Culture, and Conflict. *British Journal for the Philosophy of Science*, 67(1).
- Sterelny, K. (2017). Cultural evolution in California and Paris. *Studies in History and Philosophy of Biol & Biomed Sci*, (62), 42-50.
- Sterelny, K., Smith, K. C., & Dickison, M. (1996). The extended replicator. *Biology* and Philosophy, 11(3), 377–403.
- Stout, D. (2002). Skill and Cognition in Stone Tool Production: An Ethnographic Case Study from Irian Jaya. *Current Anthropology*, *43*(5), 693–722.
- Stout, D. (2011). Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1050–1059.
- Sugiyama, Y. (1985). The brush-stick of chimpanzees found in south-west Cameroon and their cultural characteristics. *Primates*, *26*(4), 361–374.
- Sutton, J. (2006). Distributed cognition: Domains and dimensions. *Pragmatics & Cognition*, 14(2), 235–247.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: on the

evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1528), 2405–2415.

- Terkel, J. (1995). Cultural transmission in the black rat: pine cone feeding. Advances in the Study of Behavior, 24, 119–154.
- Thorndike, E. L. (1898). Animal Intelligence: An Experimental Study of the Associative Processes in Animals. *Psychological Monographs: General and Applied*, *2*(4), i-109.

Thorndike, E. L. (1911). Animal intelligence: Experimental studies. Macmillan.

- Thornton, A., & Clutton-Brock, T. (2011). Social learning and the development of individual and group behaviour in mammal societies. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1567), 978–987.
- Tomasello, M. (1990). Cultural transmission in the tool use and communicatory signaling of chimpanzees? In S. Taylor Parker & K. R. Gibson (Eds.), Language and Intelligence in Monkeys and Apes: Comparative Developmental Perspectives. Cambridge University Press.
- Tomasello, M. (1996). Do apes ape? *Social Learning in Animals: The Roots of Culture*, 319–346.
- Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology*, (28), 509–529.
- Tomasello, M. (2009). Why We Cooperate. Cambridge, MA: MIT press.
- Tomasello, M. (2014). A natural history of human thinking. Harvard University Press.
- Tomasello, M., & Carpenter, M. (2005). Intention Reading and Imitative learning.
 In N. Chater & S. Hurley (Eds.), *Perspectives on Imitation* (Vol. II).
 Cambridge, MA: MIT Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and Brain Sciences*, *28*(5), 675-91-735.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural Learning.

Behavioral and Brain Sciences, (16), 495–552.

- Umiltà, M. A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., Gallese, V., Rizzolatti, G. (2008). When pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences*, *105*(6), 2209–2213.
- Urban, P. (2014). Joint attention and anthropological difference. *Environmental Philosophy*, *11*(1), 59–70.
- Vaesen, K. (2012). Cumulative cultural evolution and demography. *PLoS ONE*, 7(7).
- Vaesen, K., Collard, M., Cosgrove, R., & Roebroeks, W. (2016). Population size does not explain past changes in cultural complexity. *Proceedings of the National Academy of Sciences*, 113(16), E2241–E2247.
- van Baaren, R., Janssen, L., Chartrand, T. L., & Dijksterhuis, A. A. (2009). Where is the love? The social aspects of mimicry. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2381–2389.
- van Schaik, C. P. (2003). Orangutan cultures and the comparative study of culture. *American Journal of Physical Anthropology*, *102*(2003), 214.
- Voelkl, B., & Huber, L. (2000). True imitation in marmosets. *Animal Behaviour*, *60*(2), 195–202.
- Voelkl, B., & Huber, L. (2007). Imitation as Faithful Copying of a Novel Technique in Marmoset Monkeys. *PLoS ONE*, *2*(7).
- Wadley, L. (2010). Compound-Adhesive Manufacture as a Behavioral Proxy for Complex Cognition in the Middle Stone Age. *Current Anthropology*, *51*(s1), S111–S119.
- Wadley, L. (2016). Technological transformations imply cultural transformations and complex cognition. In *The nature of culture* (pp. 57–63). Springer.
- Wasielewski, H. (2014). Imitation Is Necessary for Cumulative Cultural Evolution in an Unfamiliar, Opaque Task. *Human Nature*, 25(1), 161–179.

Whiten, A. (1998). Imitation of the sequential structure of actions by chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, 112(3),

270-281.

- Whiten, A., & Ham, R. (1992). On the Nature and Evolution of Imitation in the Animal Kingdom: Reappraisal of a century of research. In J. S. R. Ed. by P. J.
 B. Slater C. Beer & M. Milinski, (Ed.), Advances in th Study of Behavior. New York: Academic Press.
- Whiten, A., & Mesoudi, A. (2008). Establishing an experimental science of culture: animal social diffusion experiments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509), 3477–3488.
- Whiten, A., & Van Schaik, C. P. (2007). The evolution of animal "cultures" and social intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 603–620.
- Whiten, A., Custance, D. M., Gomez, J., Teixidor, P., & Bard, K. (1996). Imitative learning of artificial fruit processing in children (Homo sapiens) and chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, *110*(1), 3.
- Whiten, A., Hinde, R. a, Laland, K. N., & Stringer, C. B. (2011). Culture evolves.
 Philosophical Transactions of the Royal Society of London. Series B,
 Biological Sciences, 366(1567), 938–48.
- Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009).
 Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1528), 2417–2428.
- Wilkins, J., Schoville, B. J., Brown, K. S., & Chazan, M. (2012). Evidence for early hafted hunting technology. *Science*, *338*(6109), 942–946.
- Williams, T. M., Fuiman, L. A., Kendall, T., Berry, P., Richter, B., Noren, S. R., ... Stamper, A. M. (2015). Exercise at depth alters bradycardia and incidence of cardiac anomalies in deep-diving marine mammals. *Nature Communications*, 6.
- Wimsatt, W. C. (1999). Genes, memes, and cultural heredity. *Biology and Philosophy*, 14(2), 279–310.

Wimsatt, W. C., & Griesemer, J. R. (2007). Reproduction Entrenchments to
 Scaffold Culture: The Central Role of Development in Cultural Evolution.
 Integrating Evolution and Development: From Theory to Practice, 227–324.

 Wimsatt, W. C. & Schank, J. C. (2004). Generative Entrenchment, Modularity and Evolvability: When Genic Selection Meets the Whole Organism 1.
 Modularity in Development and Evolution, 359–94.

Wood, D. (1988). How children think and learn. London: Basil Blackwell.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proceedings of the Sixth International Congress of Genetics (Vol. 1).

Yamamoto, S., Humle, T., & Tanaka, M. (2013). Basis for cumulative cultural evolution in chimpanzees: social learning of a more efficient tool-use technique. *PLoS ONE*, 8(1), e55768.

Zajonc, R. B. (1965). *Social facilitation*. Research Center for Group Dynamics, Institute for Social Research, University of Michigan Ann Arbor.

- Zentall, T. R. (1996). An analysis of imitative learning in animals. In C. M. Heyes &
 B. G. Galef (Eds.), *Social Learning in Animals* (pp. 221–243). San Diego:
 Academic Press.
- Zentall, T. R. (2001). Imitation in animal: evidence, function, and mechanisms. *Cybernetics and Systems*, *32*(1–2), 53–96.

Zentall, T. R. (2006). Imitation: definitions, evidence, and mechanisms. *Animal Cognition*, *9*, 335–353.

Zentall, T. R., & Galef, B. J. (1988). Social Learning: Psychological and biological perspectives. Hillsdale, New Jersey.

Zentall, T., Sutton, J., & Sherburne, L. (1996). True imitative learning in pigeons. *Psychological Science*, 7(6), 343–346.

Zohar, O., & Terkel, J. (1991). Acquisition of Pine Cone Stripping Behaviour in Black Rats (Rattus rattus). *International Journal of Comparative Psychology*, 5(1).