Annual Review of Earth and Planetary Sciences

The Late Heavy Bombardment

William F. Bottke\(^1\) and Marc D. Norman\(^2\)

\(^1\)Southwest Research Institute, Boulder, Colorado 80302; email: bottke@boulder.swri.edu
\(^2\)Research School of Earth Sciences, The Australian National University, Canberra ACT 0200, Australia; email: marc.norman@anu.edu.au

**Abstract**

Heavily cratere surfaces on the Moon, Mars, and Mercury show that the terrestrial planets were battered by an intense bombardment during their first billion years or more, but the timing, sources, and dynamical implications of these impacts are controversial. The Late Heavy Bombardment refers to impact events that occurred after stabilization of the planetary lithospheres such that they could be preserved as craters and basins. Lunar melt rocks and meteorite shock ages point toward a discrete episode of elevated impact flux between \(\sim 3.5\) and \(\sim 4.0–4.2\) Ga, and a relative quiescence between \(\sim 4.0–4.2\) and \(\sim 4.4\) Ga. Evidence from Precambrian impact spherule layers suggests that a long-lived tail of terrestrial impactors lasted to \(\sim 2.0–2.5\) Ga. Dynamical models that include populations residual from primary accretion and destabilized by giant planet migration can potentially account for the available observations, although all have pros and cons. The most parsimonious solution to match constraints is a hybrid model with discrete early, post-accretion and later, planetary instability–driven populations of impactors.

**Keywords**

the Moon, asteroid belt, Mars, Apollo, solar system formation, impact cratering
1. INTRODUCTION

The Late Heavy Bombardment (LHB) (Wetherill 1975) has been a central concept in planetary sciences since the 1960s, following detailed geological observations of the Moon and the discovery of petrological and geochemical evidence for intense shock metamorphism at $\sim 3.9$ Ga in many of the Apollo missions’ lunar samples (Stöffler et al. 2006). The heavily cratered terrains of the Moon (Figures 1 and 2) and other bodies such as Mercury, Mars, and Callisto provide clear physical evidence for an elevated flux of impactors across the Solar System that continued for several hundred million years after the initial accretion and differentiation of the terrestrial planets. The timing, sources, and dynamical mechanisms responsible for this bombardment, however, have been controversial and intensely debated for decades.

The LHB literature is so vast that no single article can comprehensively summarize it all. For this reason, and for readability, we focus on reevaluating several key constraints, examining new ones that have only recently become known [e.g., those from NASA’s Lunar Reconnaissance Orbiter and Gravity Recovery and Interior Laboratory (GRAIL) missions, and new geologic maps of Mars from the analysis of images taken by NASA’s Mars Reconnaissance Orbiter], and briefly summarizing the current status of dynamical models that can potentially explain early bombardment constraints. For the interested reader, considerable information can also be found in reviews provided by Ryder et al. (2000), Hartmann et al. (2000), Levison et al. (2001), Stöffler et al. (2006), Bottke et al. (2007), Fernandes et al. (2013), Fassett & Minton (2013), and Strom et al. (2015), among others. A wealth of information can also be found in Wilhelms’s (1987) classic, “The Geologic History of the Moon.”

Figure 1
A lunar map showing some of the key lunar basins, as well as the locations of the Apollo (blue stars) and Luna (green stars) landing sites where samples were returned. Basin diameters are from Neumann et al. (2015).
Historically, a major outstanding question is whether the LHB occurred as a single, short-lived episode of intense cratering at $3.9 \pm 0.1$ Ga (the Terminal Cataclysm; Tera et al. 1974) or whether it was spread over a longer interval possibly related to planetary accretion (Neukum et al. 2001, Morbidelli et al. 2012). The concept of a short-lived pulse of large impactors rampaging across the inner Solar System at $\sim 3.9$ Ga was developed by Grenville Turner (Turner 1979) and later by Graham Ryder (Ryder 1990; see Ryder et al. 2000). Following Tera et al. (1974), they proposed that a Terminal Cataclysm produced at least 15 of the major nearside lunar basins ($D > 300$ km) within an interval of $\sim 100$–$200$ Ma (see also Jessberger et al. 1974, Wilhelms 1987). Several recent studies have considered the dynamical (e.g., Levison et al. 2001, Gomes et al. 2005, Bottke et al. 2012) and astrobiological (e.g., Abramov & Mojzsis 2009) implications of this idea, even extending it to exoplanetary systems (e.g., Lisse et al. 2012). The assumption of a narrow range of ages of lunar basins has been used to calibrate the absolute ages of geological systems on other planets (Neukum et al. 2001), and the intriguing coincidence that the earliest preserved continental crust and geochemical signatures of life on Earth are also $\sim 3.9$ Ga has raised the possibility of a genetic link between the LHB and the tectonic, climatic, and biochemical evolution of the Archean/Hadean Earth (Glikson 2001, Abramov & Mojzsis 2009, Marchi et al. 2014, Nutman et al. 2016).
Alternatively, the narrow range of lunar impact ages might be due to comprehensive destruction or resurfacing of the crust toward the end of an extended period of late accretion, i.e., a stonewall effect (Hartmann et al. 2000). A prolonged episode of late bombardment might imply more benign environmental conditions on the terrestrial planets (Ryder 2002), although less frequent impacts could still have a significant influence on early atmospheres, hydrologic cycles, and biological evolution by imposing major but relatively short-lived excursions on early environmental systems (Zahnle et al. 2007). A prolonged, lower intensity bombardment may have been less likely to influence the tectonic evolution of Earth and other terrestrial planets, in which case the similarity in ages proposed for the cataclysm and observed for the oldest continental crust is likely coincidental.

The timing and duration of the early bombardment also have significant implications for understanding planetary dynamics. For example, the so-called Nice model, as presented by Gomes et al. (2005), links the LHB to the late migration of the giant planets accompanied by a scattering of comets and asteroids from stable reservoirs (see Section 3). More recently, Bottke et al. (2012) and Morbidelli et al. (2012) proposed a combination of residual planetesimals left over from accretion and asteroids escaping from the innermost region of the primordial asteroid belt to produce a late uptick in impactors combined with an extended period of heavy bombardment. The timing and sources of impacting planetesimals provide a key test for, and constraints on, the dynamical models.

In the following sections, we examine the constraints and models to see whether it is possible to distinguish between these alternatives. Doing so would strongly contribute toward a better understanding of planetary dynamics, early terrestrial environments, and biogeochemical evolution.

2. CONSTRAINTS ON EARLY BOMBARDMENT

2.1. Lunar Samples and the Ages of Impact Basins

The Moon has the most clear and complete history of early bombardment that still exists in the inner Solar System. The difficult question to address is how to properly interpret this record while also accounting for selection biases. Here, we take a new look at this long-standing problem.

2.1.1. Evidence for large impacts between 3.7 and 4.0 billion years ago. One of the unexpected results from early studies of lunar samples was the extent of impact metamorphism in samples from the Moon’s feldspathic crust. Of particular significance was the extent of textural and geochemical reworking of the primary igneous crust, and the identification of coarse-grained crystalline melt rocks that are texturally similar to terrestrial basalts but were produced by large impacts. The crystalline melt rocks are potentially the best samples for constraining absolute ages of large impact events associated with the LHB, but complications such as the presence of un-equilibrated relict clasts and the disturbance of the isotopic systems by subsequent thermal events present a significant challenge to obtaining reliable ages, especially using the commonly applied $^{40}$Ar-$^{39}$Ar system (Jessberger et al. 1974, Fernandes et al. 2013).

Even more remarkable was the pronounced clustering of crystallization ages of the lunar impact melt rocks between 3.75 and 3.95 Ga (Bogard 1995, Stößler et al. 2006). This relatively narrow range of ages corresponds to an episode of intense crustal metamorphism defined by $^{40}$Ar-$^{39}$Ar ages and U–Pb isotopic compositions of brecciated lunar anorthosites, which are thought to represent primary crust formed during igneous differentiation of the Moon (Tera et al. 1974, Premo et al. 1999). This led Tera et al. (1974) to infer a Terminal Cataclysm that they associated with the formation of several basins over a relatively brief interval of time (~200 Ma). In contrast, Schaeffer & Schaeffer (1977) suggested that the peak in $^{40}$Ar-$^{39}$Ar ages reflects the dominance
of ejecta from the Imbrium basin at all of the Apollo landing sites. A similar debate continues essentially until the present.

Critical to understanding the significance of the age distribution of lunar melt rocks is unraveling how many discrete impact events are represented, and whether these events can be linked to any of the lunar basins. Early work recognized the possibility of multiple impact events closely spaced in time (Jessberger et al. 1974, Schaeffer & Schaeffer 1977, Vaniman & Papike 1980, Spudis 1984, Korotev 1994), but the compositional variability of lunar melt rocks, the relatively large uncertainties on individual ages, and the lack of a clear geological relationship between individual samples and impact structures hampered the resolution of discrete events (Haskin 1998, Haskin et al. 1998).

In a more recent study, Norman et al. (2006) proposed at least four discrete impact events ranging in age from 3.75 to 3.95 Ga based on groupings of ages, compositions, and textural characteristics of crystalline melt rocks from the Apollo 16 site (Figures 1 and 2). The range of Rb-Sr isochron ages and initial $^{87}$Sr/$^{86}$Sr ratios of crystalline lunar melt rocks (Papanastassiou & Wasserburg 1972, Taylor et al. 1991, Bogard 1995, Stöffler et al. 2006, Nyquist et al. 2011) are also consistent with multiple impacts during that time. Crucially, the pronounced clustering of lunar melt rock ages in the interval typically associated with the Terminal Cataclysm is apparent in diverse isotopic systems, including Rb-Sr, Sm-Nd, U-Pb, and $^{40}$Ar-$^{39}$Ar. Although $^{40}$Ar-$^{39}$Ar data can be subject to post-crystallization disturbance and other uncertainties (Boehnke & Harrison 2016), the excellent agreement obtained by multiple techniques provides confidence that the ages are meaningful. A straightforward interpretation of these data is that numerous impacts large enough to generate abundant crystalline melt rocks (requiring impact structures tens to hundreds of kilometers in diameter) occurred within the interval from 3.7 to 4.0 Ga, and that such events were much less common on the Moon after that time.

Lunar meteorites provide another source of information about impact history from locations potentially far removed from Imbrium. Many lunar meteorites are regolith breccias that carry clasts of impact melt rocks (a breccia is a rock composed of fragments or clasts of preexisting rocks and minerals in a finer-grained matrix). Although relatively few crystallization ages have been determined for these melt rock clasts, their distribution is broadly similar to that inferred from the Apollo samples and crater density studies of the lunar surface, with the oldest apparent ages $\sim$4.2 Ga, a peak at $\sim$3.7 Ga, and a declining number of ages to $\sim$2.5 Ga (Cohen et al. 2005, Joy & Arai 2013). As many of the feldspathic lunar meteorites are regolith breccias, the longer tail to younger ages in the meteorite clasts compared to the Apollo melt rocks may reflect smaller or more localized impact events (Joy & Arai 2013).

### 2.1.2. Ages of lunar basins

From this broad viewpoint, we now examine the links between lunar samples and the ages of multiring impact basins ($D > 300$ km), all of which are older than $\sim$3.7 Ga. At least 40 and possibly up to 90 lunar basins have been recognized or proposed (Wilhelms 1987, Frey 2011, Fassett et al. 2012, Neumann et al. 2015). Geological relationships have been used to construct a relative time sequence of the lunar basins, although the stratigraphic position of several basins remains unclear (Wilhelms 1987, Fassett et al. 2012).

Even more uncertain, however, are the absolute ages of lunar basins, despite increasingly precise measurements of the isotopic ages of lunar samples and increasingly detailed geological studies of the lunar surface using high-resolution imaging. This uncertainty largely reflects our inability to definitively link individual lunar samples with specific basins or craters. Stöffler et al. (2006) provided a comprehensive summary of the radiometric ages of lunar samples available at that time, and they proposed correlations with some key nearside basins such as Imbrium, Serenitatis, Crisium, and Nectaris (Figures 1 and 2). The youngest basins (such as Imbrium and Orientale) are clearly $\sim$3.7 to $\sim$3.9 Ga. Imbrium has perhaps the most clearly defined age of 3.91–3.94 Ga.
Gnos et al. 2004, Nemchin et al. 2009, Liu et al. 2012, Merle et al. 2014, Snape et al. 2016) based on U-Pb isotopic ages of zircons and apatites in lunar breccias, with high concentrations of Th and other incompatible trace elements (ITEs) similar to the region of the Moon where the Imbrium basin is located (the Procellarum-KREEP Terrain; Jolliff et al. 2000). Beyond that broad constraint, there is continuing uncertainty and debate.

2.1.3. Interpretation of ages at Apollo 14 and 17 landing sites. Another long-standing debate with significant implications for the Terminal Cataclysm hypothesis concerns the geology of the Apollo landing sites (Figure 1) and the diverse suite of impact breccias and melt rocks collected there. In the 1980s, lunar geologists came to a consensus about the genetic relationship of ejecta deposits sampled by the Apollo missions to the Imbrium, Serenitatis, and Nectaris basins. According to this consensus, the Fra Mauro formation sampled by Apollo 14 was primary Imbrium ejecta, the massif unit sampled by Apollo 17 was Serenitatis impact melt, and the Cayley and Descartes units at the Apollo 16 site in the central nearside highlands were most likely deposits from Imbrium and Nectaris, respectively. Breccias collected at these sites were interpreted in this context (Stößler et al. 2006). Although no ejecta deposits were sampled specifically by the Apollo 15 mission, two unusual samples collected there were considered to be prime candidates for melt rocks formed by the Imbrium basin-forming impact.

Today, most of these relationships have been questioned and are under active debate. For example, although the regional physiographic characteristics of the Fra Mauro formation (Apollo 14) were clearly produced by Imbrium, it has so far proven impossible to convincingly distinguish in the available samples primary Imbrium ejecta from products of secondary and local craters (Spudis 1993, Merle et al. 2014). Hence, although the ages of these breccias and their components can be measured to very high precision, linking these data to specific basins or craters remains speculative (Nemchin et al. 2009, Merle et al. 2014, Snape et al. 2016).

In contrast, Apollo 17 was intended to sample the Serenitatis basin, which is stratigraphically older than Imbrium based on cross-cutting relationships observed in orbital images, although the exact position of Serenitatis in the sequence of lunar basins is somewhat uncertain (Spudis et al. 2011, Fassett et al. 2012) (Figure 1). Attempts to date Apollo 17 breccias using $^{40}$Ar-$^{39}$Ar returned a slightly older age compared to Apollo 14 and 15 breccias considered as the best candidates for Imbrium ejecta (Dalrymple & Ryder 1993, 1996). This led to the conclusion that Serenitatis is 20–40 Ma older than Imbrium, thereby supporting the idea that multiple basins formed within a narrow interval of time, i.e., a Terminal Cataclysm.

Those isotopic dating studies were state-of-the-art at the time, but the Ar release spectra were complex and the inferred age of Serenitatis depended on a careful selection of the data. In addition, recent reinterpretations of stratigraphic relationships at the Apollo 17 site using new, high-resolution imaging from the Lunar Reconnaissance Orbiter have cast doubt on the geological setting of the site by linking exposed deposits more directly with Imbrium rather than Serenitatis (Spudis et al. 2011, Fassett et al. 2012). An alternative interpretation proposed by Haskin et al. (1998) was that the Apollo 17 melt rocks contained excess Ar probably derived from entrained clasts (see also Jessberger et al. 1977). We conclude that the currently accepted textbook ages of the Serenitatis basin (Stößler et al. 2006) are uncertain.

2.1.4. Interpretation of ages at Apollo 16 landing site. Even more challenging is interpreting the diverse suite of impact breccias and melt rocks collected by Apollo 16 (Figure 1). This mission targeted two regionally significant geological formations with distinct physiography, the Cayley Plains and the Descartes Hills. Although pre-mission planning expected these units to represent different styles of highlands volcanism, the origin of these units as impact deposits was quickly...
realized (Muehlberger et al. 1980; see also Hiesinger & Head 2006). There is now general, but not universal, agreement among lunar geologists that the Cayley Plains unit is correlative with the Fra Mauro formation and that it was deposited as, or extensively reworked by, Imbrium ejecta, possibly followed by later resurfacing and addition of material from younger craters (Spudis 1993, Korotev 1994, Neukum et al. 2001, Stöffler et al. 2006, Joy et al. 2011). The origin of the Descartes unit has been more contentious. Based on geological relationships observed from orbit, lunar geologists have proposed that its samples originated as ejecta from either Imbrium or the older Nectaris basin (Muehlberger et al. 1980; see also Hiesinger & Head 2006). The key question is whether Nectaris was dated by Apollo 16 samples; Nectaris is stratigraphically older than about a third of all of the lunar basins, so its age would provide an important constraint on the timing of the LHB.

A diverse suite of crystalline melt rocks was collected as float from the surface of the Cayley unit at the Apollo 16 site (Vaniman & Papike 1980, Norman et al. 2006). Because the survival times of meter-size boulders on the lunar surface are relatively short, these samples were probably emplaced on the lunar surface by relatively small, recent craters and their primary geological context is not well constrained. Although it is therefore difficult to link individual samples with specific geological formations, much less any particular basin or crater, the best candidates for primary Imbrium ejecta may be a set of melt rocks with high concentrations of ITEs that imply a source in the Procellarum-KREEP Terrain (Group 1 of Korotev 1994; Norman et al. 2006, 2010). Crystallization ages of this group of Apollo 16 melt rocks overlap within uncertainty with the zircon and apatite ages of impact breccias from the Apollo 14 site (3.92–3.94 Ga after correction for laboratory constants; see Stöffler et al. 2006, Nemchin et al. 2009, Norman et al. 2010, Merle et al. 2014, Snape et al. 2016). We interpret this as the best estimate of the age of the Imbrium basin from currently available data. Other groups of Apollo 16 melt rocks have younger ages and compositions that are both more aluminous and lower in ITE concentrations, which suggests they formed by smaller, post-Imbrium craters ($D < 300$ km) in crustal terrains more proximal to the Apollo 16 site.

More contentious is the origin and geological significance of the Descartes Hills, which were sampled by Apollo 16 at North Ray crater. Early geological studies of the Descartes formation using images obtained from lunar orbit proposed that the Descartes and Cayley units were coeval and represent different facies of Imbrium ejecta deposits (Muehlberger et al. 1980). Based on remote sensing data and the samples collected from the rim of North Ray, however, the Descartes unit consists predominantly of low-grade, fragmental, or shock-lithified breccias with highly aluminous compositions and low concentrations of ITEs. These compositional characteristics are more consistent with a derivation from the type of feldspathic crust surrounding the Nectaris basin (also referred to as the Felspathic Highlands Terrain; Jolliff et al. 2000) rather than from the Procellarum-KREEP Terrain in which Imbrium is located. James (1981) and others reinterpreted the Descartes unit as Nectaris ejecta and proposed an age for Nectaris of 3.92 Ga based on $^{40}$Ar-$^{39}$Ar data from a few clasts of metamorphosed breccias extracted from the regolith around North Ray crater. This “age of Nectaris” provided strong support for the Terminal Cataclysm hypothesis in which many, if not all, of the lunar basins formed in a single episode of intense bombardment (Ryder 2002). Subsequent detailed study of the Descartes breccias, however, showed that the youngest age population is coeval with the ITE-enriched group of Apollo 16 melt rocks discussed above, and that some lithic fragments carried by the Descartes breccias are petrographically similar to basalt collected at Apollo 14 (see Norman et al. 2010, and references therein). These observations appear more consistent with a genetic relationship of the Descartes breccias to Imbrium rather than Nectaris.

Another intriguing aspect of the Descartes breccias is the presence of even older components with apparent ages $> 4.1$ Ga (e.g., Delano & Bence 1977, Schaeffer & Schaeffer 1977, Maurer et al. 1978, Norman et al. 2010, Shuster et al. 2010, Fernandes et al. 2013). Although the petrologic
affinities of these pre-Imbrian component(s) were not always well established, they seem to be associated with the Descartes breccias. Schaeffer & Husain (1974) proposed that these ages date Nectaris at 4.25 ± 0.05 Ga, whereas Maurer et al. (1978) considered the ≥4.1 Ga components in the Descartes breccias to be related to sub-basin-size craters (D ∼ 100–300 km). Some metamorphic lunar breccias and impact-deformed lunar zircons also yield such ages (Hudgins et al. 2008, 2011; Grange et al. 2013), but the significance of these data for lunar impact history is unclear because the scale of events represented by these materials is not well defined. Fernandes et al. (2013) have provided a comprehensive discussion of these issues.

A new perspective on the significance of pre-Imbrian components in the Descartes breccias was presented by Norman & Nemchin (2014) and Norman et al. (2016). They showed that a coarse-grained, igneous-textured lithic clast from the Descartes breccias formed in a basin-scale impact event, and they dated that event at 4.22 ± 0.01 Ga. They also argued that it crystallized in a central melt sheet and that the impact that formed it occurred in the Procellarum-KREEP Terrain rather than in the feldspathic highlands. These petrological and geochemical characteristics imply transport of the clast to the Apollo 16 site as Imbrium ejecta rather than Nectaris ejecta. Unfortunately, this 4.22 Ga melt rock cannot be related directly to any particular basin. Speculative possibilities include basins near Imbrium (e.g., Serenitatis) or an unrecognized structure that was destroyed by Imbrium. Of course, there might be other lunar basins, including Nectaris, that are also 4.2 Ga (or any age older than Imbrium), but samples that can be related convincingly to any of those basins have not been recognized so far.

In summary, current data provide compelling evidence that the LHB extended back in time to at least 4.2 Ga and possibly before. The LHB is real. The strongest version of the Terminal Cataclysm hypothesis in which all of the lunar basins formed within a brief interval (≤200 Ma), however, can be excluded as a viable hypothesis. The nearside region of the Moon seems to have been comprehensively resurfaced by Imbrium ejecta, and this has biased our view of the Moon based on the Apollo samples. Compositional variations within Imbrium ejecta appear to reflect complexities in radial ejecta distributions rather than separate impact events. The lack of absolute ages, especially for the older lunar basins, and solid constraints on the mass versus time flux of impactors across the inner Solar System is a significant impediment to resolving the nature of the LHB.

2.2. Small Lunar Craters and Crater Chronology

Impact craters are ubiquitous across the Solar System and have long been used as a metric for assigning relative ages (surfaces with higher spatial densities of craters are older than surfaces with lower crater densities). These densities give us a sense of the earliest lunar bombardment. A major goal of the Apollo and Luna missions (Figure 1) was to calibrate this relative scale to an absolute one: Ideally, crater densities were measured on geologic units from which dated samples were collected, yielding the fundamental chronologic relationship between crater density and absolute age (Wilhelms 1987, Hartmann et al. 2000, Neukum et al. 2001, Stöffler et al. 2006, Robbins 2014).

This led to curves like those shown in Figure 3. By identifying 0.3 < D < 3 km craters on dated terrains, it was possible to derive N(1), the expected density of D > 1 km craters per square kilometer on the Moon. These flux curves can be scaled to larger craters by assuming the impact size–frequency distribution (SFD) stays the same shape over time. According to this logic, even though D > 1 km craters may have potentially reached saturation in some ancient lunar regions, the effective N(1) for that region can be obtained by calculating N(20), N(50), etc. and rescaling, provided that sufficient crater statistics are available. At that point, dynamical models of small body impact rates on other inner Solar System planets can be used to determine absolute ages of geologic units on their surfaces, using the lunar data for calibration.
Figure 3

Different estimated lunar chronologies. The x-axis is past time in units of billions of years from today, and the y-axis is $N(1)$, the spatial density of diameter $D > 1$ km craters per square kilometer. The curves are based on the $N(1)$ values found on Apollo and Luna terrains where samples have been collected and dated. This oft-used chronology from Neukum et al. (2001) can be considered a proxy for a declining bombardment scenario (red curve). They assume that the Apollo 14 and 16 sites with the highest crater spatial densities are nearly 4.1 Ga. Robbins (2014) provided an alternative that is a reasonable proxy for a Terminal Cataclysm scenario (blue curve). His curve is different because he recounted craters at the Apollo and Luna sites using data from NASA’s Lunar Reconnaissance Orbiter, and he assumed that the Apollo 14 and 16 sites have ages near 3.9 Ga. The absolute ages of the Apollo 14 and 16 sites are unknown. A hybrid model, a sawtooth scenario (dark yellow curve), was provided by Morbidelli et al. (2012). They assumed two phases of bombardment: Leftover planetesimals strike earliest, whereas late giant planet migration starts a second wave of impactors near 4.1 Ga.

We refer to the crater chronology curve from Neukum et al. (2001) in Figure 3 as a declining bombardment. This curve was originally generated on the basis of geological work on lunar impact chronology done in the 1970s using images obtained by Lunar Orbiter and Apollo photographs. These pioneering missions obtained high-quality images but their coverage of the lunar surface was limited. High-resolution images from the Lunar Reconnaissance Orbiter Wide-Angle Camera, however, have revolutionized our ability to understand the Apollo and Luna terrains. A remapping of all 11 Apollo/Luna landing sites by Robbins (2014) led to crater counts different than those found previously (Figure 3). This suggests that long-adopted lunar impact chronologies may have substantial inaccuracies, even without consideration of the new interpretation of the sample ages found at the Apollo sites (see Section 2.1).

Despite these limitations, some age insights can still be gleaned from the relative spatial densities of craters superposed on the largest lunar basins. For example, the ratios of the spatial densities of $20 < D < 100$ km craters on the basins Nectaris/Imbrium and Nectaris/Orientale, when compared as cumulative SFDs, are $\sim 4$ to 5 and $\sim 7$ to 10, respectively (Fassett et al. 2012). With the assumption that the Neukum et al. (2001) curve in Figure 3 represents the shape of the impact
flux curve, and that Imbrium’s age is 3.9 Ga, the Neukum curve would predict that the age of Orientale is ∼3.7 Ga and Nectaris is ∼4.1 Ga. Similarly, the oldest lunar terrains show crater spatial densities that are a factor of 2 or so higher than Nectaris (Marchi et al. 2012). This would suggest the most ancient lunar farside terrains are only ∼4.2 Ga, a surprisingly young age given that the lunar crust may be ≥4.35 Ga (Borg et al. 2015).

A possible way to increase the ages of the oldest cratered terrains would be to assume that the impact flux changed in some manner >3.9 Ga. In Figure 3, we show a possible impact flux curve from Morbidelli et al. (2012). Here it is assumed that there were two impact populations: an early one produced by planetesimals left over from primary accretion and a second one produced by late migration of the giant planets that took place at 4.1 Ga. This hybrid model, which combines aspects of declining bombardment and Terminal Cataclysm models, would potentially modify the ages of the oldest lunar terrains to ∼4.35–4.5 Ga. Morbidelli et al. (2012) referred to this kind of pattern as a sawtooth, based on what the impact flux curve looks like on a differential plot. A Terminal Cataclysm is an extreme version of the sawtooth, with the tooth becoming a spike. These various scenarios are discussed below.

### 2.3. Medium-Sized Lunar Craters

There are numerous assertions and controversies surrounding the interpretation of ancient 20 < D < 100 km craters on the Moon. The arguments have two basic flavors: (a) The shape of the lunar crater SFD either stayed about the same or changed substantially as the basin-forming era wore on, and (b) these craters either were or were not made by a projectile population with an SFD like the present-day main asteroid belt. The details are too extensive to discuss here, but a short list of relevant articles on these topics would include Strom et al. (2005, 2015), Ćuk et al. (2010), Head et al. (2010), Fassett et al. (2012), Marchi et al. (2012), Minton et al. (2015a,b), Neumann et al. (2015), Johnson et al. (2016), Bottke et al. (2016), and Fassett (2016).

Although this debate is fascinating, there are issues that need to be considered when evaluating these topics. First, craters derived from a given production population in a stochastic manner can produce a wide range of crater SFDs when the number of draws is sparse. This can be readily demonstrated using a Monte Carlo code. We suspect that the changes in crater SFDs between ancient and less ancient terrains might be explained in this manner (e.g., Head et al. 2010, Fassett et al. 2012). A related issue involves choosing the appropriate surface for crater counting, which can be difficult even after a geologic mapping campaign.

Second, there is no agreed-upon estimate of the shape of the main asteroid belt’s SFD for 1 < D < 5 km, even though these bodies produce most of the 20 < D < 100 km craters. The problem is twofold: Only a portion of all D < 5 km asteroids have been detected, and the albedos of the discovered objects in this size range have not yet been fully debiased (Masiero et al. 2013). This makes it more difficult to argue that a given crater SFD does or does not precisely fit the main belt SFD.

Third, there is no agreed upon crater scaling law formulation that one should use to make D > 20 km craters on the Moon or elsewhere (e.g., see contrasting views in Johnson et al. 2016 and Bottke et al. 2016). A critical issue here is the lack of constraints; most of the data used to derive crater scaling laws for big impacts come from either high-velocity laboratory shot experiments or relatively small nuclear blasts (Melosh 1989).

Until follow-up work settles these issues, we modestly favor the hypothesis that crater SFDs seen on ancient lunar terrains are, at the least, comparable to the wavy SFD seen in the asteroid belt, and that the shape of the projectile SFD that made most 20 km < D < 100 km craters on the Moon did not substantially change over time. We return to this topic in in Section 2.6.
2.4. The Largest Lunar Craters

It is possible that the best way to understand the earliest lunar bombardment flux is through large impact craters \((D > 100-300 \text{ km})\). They represent most of the mass of the impacting size distribution, with topographic and gravity signatures that are hard to erase by geologic processes or subsequent impacts. Geologic mapping of the Moon indicates that the oldest lunar terrains are located on the lunar farside (pre-Nectarian terrain; Wilhelms 1987, Head et al. 2010, Fassett et al. 2012, Neumann et al. 2015). This can also be seen from the spatial density of \(D > 150 \text{ km}\) craters identified by GRAIL gravity data (Figure 4) (Neumann et al. 2015). The SFD of farside \(D > 150 \text{ km}\) craters is shown in Figure 5a. We discuss this record again in Section 2.6. Historically, however, most studies of early bombardment have concentrated on understanding the crater record on the lunar nearside as this was where the best images and sample ground truth were available. Interpreting the nearside impact record, however, is surprisingly complicated.

The Procellarum region, a large quasi-circular feature that covers most of the lunar western nearside (Figure 1), was once considered to be a gigantic 3,200 km basin itself, but new data from NASA’s GRAIL mission suggest it may instead be a gigantic tectonomagmatic structure (Andrews-Hanna et al. 2014). The thin crust and high concentrations of the heat-producing elements uranium, thorium, and potassium in this region (Wieczorek et al. 2006) allowed it to heat up and stay hotter for hundreds of millions of years longer than the farside terrain, which has a much thicker crust. In addition to reprocessing of the crust by magmatism, numerical modeling indicates that transient craters formed in a hot crust will collapse more extensively.
than those formed in a cold crust (Miljković et al. 2013). This high heat flow in the Procellarum region therefore allowed substantially larger basins to form on nearside than on farside terrains, explaining an apparent asymmetry in the spatial distribution of lunar basins (Figure 4). The formation and evolution of the Procellarum region also appear to have erased the most ancient nearside basins and terrains. The paucity of lunar melt rocks >4.2 Ga could therefore be at least partially a by-product of this region’s evolution.

Although it is often an afterthought, an additional useful LHB constraint is the population of post-Orientale craters with \( D > 150 \) km. They tell us about the LHB’s tail of smaller impactors (Bottke et al. 2012). Terrains dated by Apollo samples, combined with geological mapping, indicate that four \( D > 150 \) km craters formed between \( \sim 3.0 \) and 3.7 Ga, three between 3.2 and 3.7 Ga (Iridium, Humboldt, and Tsiolkovskiy, with diameters of 260, 207, and 180 km, respectively; Wilhelms 1987), and one between \( \sim 3.0 \) and 3.5 Ga (Hausen, 167 km in diameter; Wilhelms 1987, Kirchoff et al. 2013).

2.5. Constraints from Mars

One of the best geological records of early bombardment comes from Mars. Mars has four times the surface area of the Moon, and silicate differentiation apparently occurred within \( \leq 60 \) Ma of the origin of the Solar System (Borg et al. 2016). Assuming that the surface of Mars formed during initial differentiation, it would have borne witness to the primary populations of planetesimals that were accreting to form the terrestrial planets.
A long-standing controversy about Mars concerns the endogenic (i.e., tectonomagmatic) versus exogenic (i.e., impact-related) origin of its crustal dichotomy, where the rugged southern highlands stand in dramatic contrast with the relatively flat northern lowlands. Detailed geological observations of Mars from spacecraft missions coupled with dynamical simulations now provide strong support for the idea that the lowlands are the result of a massive early collision that excavated and redistributed the primordial crust from the northern hemisphere (Andrews-Hanna et al. 2008, Marinova et al. 2008, Nimmo et al. 2008). This would make the martian lowlands, now referred to as the Borealis basin, the largest known impact structure in the Solar System. The Borealis basin is estimated to be 10,600 km by 8,500 km in area, covering 40% of Mars’s surface (Figure 6). The collision that formed Borealis probably occurred during the waning stages of planetary accretion (Nimmo et al. 2008, Bottke & Andrews-Hanna 2017), and likely reset the radiogenic isotopic systems of much of the martian crust. If so, the Borealis impact is at least as old as the $\sim$4.43 Ga zircons found within the paired martian meteorites NWA 7533 and 7034 breccias that likely represent lithified regolith (Humayun et al. 2013).

Ancient martian crustal-scale basins with clear physiographic and geophysical expressions include Hellas, Isidis, and Argyre; all three are located in the southern highlands (Figure 6) and they appear to be younger than Borealis. In the northern lowlands, only the Utopia basin has an unambiguous gravity signature comparable to these basins (Thomson & Head 2001, Bottke & Andrews-Hanna 2017) (Figure 6). Given the diameters of their crustal thickness cavities, these basins are 1,940 km, 1,500 km, 780 km, and 2,200 km across, respectively. Crater retention model
ages tied to the absolute timescale established for the Moon (see Section 2.2) yield ages of \( \sim 3.8-4.1 \) Ga for the southern basins (Robbins et al. 2013), near the interval typically associated with the Terminal Cataclysm.

Many tens of additional basins \( 900 \text{ km} < D < 3,000 \text{ km} \) have been suggested based on features in topographic and crustal thickness models (Frey & Mannoia 2013). Statistical models indicate that a post-Borealis bombardment population of this magnitude would take numerous bites out of the highlands-lowlands boundary, compared to only the single observed intersection by Isidis (Bottke & Andrews-Hanna 2017). This analysis limits the number of pre-Noachian crustal-scale basins formed after Borealis to less than 12. An analysis of gravity signatures further winnows this list to those discussed above: Hellas, Utopia, Argyre, and Isidis. This could favor an impact chronology in which the mega impact responsible for the Borealis basin at \( > 4.43 \) Ga was followed by a period of relative quiescence prior to the formation of the largest Noachian basins between \( \sim 3.8 \) and 4.0 Ga (Bottke & Andrews-Hanna 2017). Mars’s remaining record of large impacts is discussed in Section 2.6.

Unfortunately, martian meteorites provide few constraints on the early impact history of Mars because most of these samples are igneous rocks that formed well after the basin-forming era had ended. The carbonates in martian meteorite ALH84001 appear to have formed \( \sim 4.0 \) Ga (Borg & Drake 2005) but the significance of this determination is unclear. Perhaps of more critical significance for this discussion may be the convergence on a primary crystallization age of \( \sim 4.1 \) Ga for ALH84001 compared to earlier indications that the rock formed at \( \sim 4.5 \) Ga (Lapen et al. 2010). If this meteorite samples the heavily cratered southern hemisphere of Mars, that surface may be much younger than is generally assumed.

### 2.6. Comparison of Large Impact Craters on Mars and the Moon

Up to now, most LHB studies have concentrated almost entirely on the Moon. Given how much we have learned about martian bombardment over the last several years, however, the time seems ripe to compare and contrast the large crater populations of both worlds. In considering Mars, we look to the 98 craters with \( D > 150 \) km that are located predominately on the ancient southern highlands of Mars (Figures 5 and 6). (Tanaka et al. 2014). They cover approximately half of Mars and were identified as a byproduct of an extensive geologic mapping effort. Buried or equivocal structures were excluded, and it can be argued the list is fairly conservative (e.g., Robbins et al. 2013). Superposed crater counts suggest most are early- to mid-Noachian in age and therefore represent some of the oldest martian terrains.

As shown in Figure 5a, the raw crater SFDs of the lunar farside and martian highlands have somewhat different shapes, but this could be because we are not examining the most densely cratered terrains. For example, young lunar farside basins like Orientale, Hertzprung, and Moscoviense (Figures 1 and 4) appear to have erased all of the ancient terrains near them. The most densely cratered lunar terrains, which cover approximately 70% of the farside, closely adhere to the pre-Nectarian terrains identified by Wilhelms (1987). Scaling the modified crater SFD for this surface area, we get a modestly closer match between Moon and Mars in Figure 5b. Both show a limited number of basins with \( D > 400-600 \) km that shift at smaller crater sizes to cumulative power-law slopes of \(-2.1\) to \(-3.1\) for the Moon and \(-2.6\) to \(-3.1\) for Mars.

Although the shapes of these crater SFDs are modestly different from one another, they are also potentially reminiscent of the main belt SFD discussed in Bottke et al. (2005) for \( 5 < D < 200 \) km asteroids (see also Morbidelli et al. 2009, Masiero et al. 2013, Bottke et al. 2015) (Figure 7). Using a collisional evolution code, Bottke et al. (2005) showed that a primordial main belt asteroid population dominated by \( D > 100 \) km bodies undergoes comminution and builds up a fragment...
Figure 7
Four snapshots from a representative run where Bottke et al. (2005) tracked the collisional evolution of the main belt size frequency distribution. The dots are the observed population. For clarity, primordial asteroids that escape from the main belt are not shown; they also contribute to comminution. The bump in the size distribution near 100 km is a by-product of accretion; most asteroids are assumed to be “born big” (Morbidelli et al. 2009). The fragment tail and bump near 3 km are produced by the nature of the asteroid disruption law. The slope of the fragment tail between 5 < D < 20 km is approximately −3 cumulative, similar to the slopes in Figure 6.

tail of D < 20–30 km bodies. The slopes achieved for those fragment tails can be near −3.0, similar to those observed (Masiero et al. 2013). Similar behavior is expected among colliding planetesimals in the terrestrial planet zone (Bottke et al. 2007) and colliding comet-like objects formed in the outer Solar System (Levison et al. 2009, Parker & New Horizons Sci. Team 2015).

Accordingly, these inferences suggest the Moon and Mars were likely bombarded by one or more collisionally evolved populations. Moreover, if we assume the shapes of the crater SFDs are directly comparable to the results from Bottke et al. (2005), the steepest portions of the crater SFDs in Figure 5 were likely made by the steepest portion of the fragment tails found in Figure 7 (i.e., projectiles between ~5 and 20–30 km), whereas the largest lunar and martian basins were made by projectiles that ranged from 30 km to a few hundred kilometers in diameter. This same relationship would explain why the crater SFDs found on the Moon and Mars for 20 km < D < 100 km craters have wavy shapes—they come from the wavy part of the asteroid/comet/planetesimal SFD produced by collision evolution that happens to be located between 1 km < D < 5 km (Strom et al. 2005).

The impact ratios between Mars and the Moon provide us with additional constraints on the LHB, although further interpretation of the data is needed. For example, accepting the values in Figure 5a,b at face value, and assuming the crater SFDs in Figure 5b are essentially the same, they suggest that the Mars to Moon impact ratio is near 1, a surprisingly low number given that the surface area of Mars is nearly four times that of the Moon. If we instead use the most densely cratered terrains on Mars from Figure 6, as we did for the Moon, we find that about half of the craters cover only 10% of the martian surface. Scaling for its entire surface, we get roughly 500 craters for Mars compared to 180 for the Moon. This yields a Mars to Moon impact ratio of ~2.8. Finally, an impactor striking Mars may yield a smaller crater than one hitting the Moon (Melosh 1989). This result is suggested in the difference in the inflection points near D ≈450 km and 650 km from Figure 5b; if the scaling laws were the same, the inflection points should be at the same diameter. Using this simple factor and applying it to compare the most densely cratered terrains on Mars and the Moon yields an impact ratio of 5.3. The dynamical interpretation of early bombardment strongly depends on which Mars to Moon impact ratio is used (1, 2.8, or 5.3) and on the assumption the crater SFDs are the same. The implications of these values are discussed in Section 4.
2.7. Constraints from Meteorites

Additional constraints come from shock ages of meteorites derived from asteroid parent bodies, with most of these based on $^{40}\text{Ar}-^{39}\text{Ar}$ data, typically of rock chips, and U-Pb isotopic compositions of phosphates, as these systems are sensitive to lower $T$ events (reviewed in Bogard 2011, Jourdan 2012, and Swindle et al. 2014). Shock degassing ages are produced by collision events on sizeable asteroids that yield high temperatures beyond the Ar loss threshold. They have already been used to successfully link disparate Solar System events—e.g., (a) two-thirds of shocked L-chondrites have $^{40}\text{Ar}-^{39}\text{Ar}$ reset ages near 470 Ma; (b) these ages coincide with the stratigraphic age ($467 \pm 2$ Ma) of the mid-Ordovician strata where abundant fossil L-chondrites, meteorite-tracing chromite grains, and iridium enrichment were found in the active marine limestone quarry in southern Sweden; and (c) the samples are likely connected with the disruption of the Gefion parent asteroid, whose dynamical-derived age is 480 ($+40, -10$) Ma (Nesvorny et al. 2009).

The collision events most likely to produce Ar loss temperatures are those where the target asteroid is hit at a velocity greater than 10 km/s (Marchi et al. 2013). Such velocities seldom occur when main belt asteroids strike one another, which probably explains the paucity of shocked rocks and impact melt in current meteorite collisions and on large asteroids observed in situ like Vesta. To reach velocities over 10 km/s, projectiles instead need sizable inclinations, which are rare in the present main belt, and/or eccentricities larger than 0.5, which take them out of the main belt entirely and onto planet-crossing orbits. Such excited impactors, however, have dynamical lifetimes of only a few to a few tens of millions of years. Accordingly, meteorite shock degassing ages may tell us about transient populations on planet-crossing orbits.

Here we focus on evidence for shock events that occurred at $>3$ Ga. A fundamental limitation of these studies is that many of the meteorites are petrologically complex. The best available samples are clast-poor impact melt rocks such as LAP 02240 and LAP 031125 (Swindle et al. 2009), but these are rare and many samples are either melt rocks with abundant relict clasts or lower-grade breccias with evidence for variable shock and fragmentation (Swindle et al. 2009, 2014; Bogard 2011). Some are unbrecciated and apparently experienced only thermal annealing after their initial formation. As a consequence, $^{40}\text{Ar}-^{39}\text{Ar}$ shock ages from chondritic and achondritic meteorites are often not as readily interpretable as they are for many of the lunar melt rocks. In many cases, the apparent age spectra obtained from step-heating experiments are complex and produce patterns that indicate incomplete resetting or other effects such as recoil, later diffusive loss, or multiple K-bearing domains.

Some reasonably consistent patterns, however, have emerged that illuminate broad aspects of the shock histories of asteroids relevant to the LHB. The age patterns can be broadly divided into three main groups: numerous samples between 4.55 and $\sim4.40$ Ga, the expected era of planet formation; relatively few samples between 4.0 and 4.4 Ga; and numerous samples between 3.4 and 4.0 Ga, the period commonly associated with elevated cratering rates on the Moon and the Terminal Cataclysm. Numerous classes of meteorites show evidence for either extended cooling or discrete thermal events on their parent bodies within $\sim50$–150 Ma after they formed. These include the H and L ordinary chondrites, enstatite chondrites, unbrecciated basaltic and cumulate eucrites, primitive achondrites such as lodranites and acapulcoites, and silicate inclusions in some types of iron meteorites. Clearly, this reflects a widespread event or set of conditions that affected various chondritic and differentiated asteroids. Although slow cooling or metamorphism of the parent bodies might account for some of these results, there is also petrographic evidence in some meteorites for impact melting during this interval (e.g., PAT 91501, Shaw, and MIL 05029; see Bogard 2011).

Some classes of meteorites also show clear petrographic and isotopic evidence for impact events that occurred several hundred million years after initial accretion. For example, H-chondrites show...
a prominent group of reported $^{40}$Ar-$^{39}$Ar ages between $\sim$3.5 and 4.0 Ga, with the clast-poor impact melt rocks LAP 02240 and LAP 031125 yielding especially well-defined plateau ages of $3.939 \pm 0.062$ Ga and $3.942 \pm 0.023$ Ga, respectively (Swindle et al. 2009, 2014). Eucrites, believed to come from the asteroid Vesta, have also yielded numerous Ar ages between 3.4 and 4.1 Ga, with groups of reliable ages at $\sim$3.5 and 3.8–4.0 Ga, and few such ages between 4.1 and 4.5 Ga (Bogard 2011, Kennedy et al. 2013). Mesosiderites commonly yield Ar ages of 3.8–4.1 Ga, with a mean age of $3.94 \pm 0.1$ Ga for the 19 samples considered by Bogard (2011), although the very slow cooling experienced by these meteorites following disruption and re-accretion of their parent body complicates the interpretation of these data.

The story that emerges from shock age distributions of asteroid-derived meteorites seems to be that primary accretion lasted $\leq$150 Ma, and that associated collisions were widespread and affected many different types of bodies. This was followed by a period of relatively few collisions capable of imprinting geochemical signatures on the commonly applied radiogenic isotope systems before another episode of impacts left a distinctive fingerprint at $\sim$3.5–4.0 Ga. Although this younger episode also affected a diverse suite of meteorite parent bodies, its effects were distributed heterogeneously, and it does not seem to have been either as intense or as widespread as the tail that followed initial accretion.

Overall, the pattern of shock ages obtained from asteroid-derived meteorites appears remarkably consistent with that observed in the lunar samples. Although this appears to be inconsistent with the strong version of the Terminal Cataclysm, it also seems to require an episode of increased late bombardment that is distinct from the tail of primary accretion.

### 2.8. Terrestrial Constraints on Early Bombardment

The earliest history of Earth is poorly understood because few rocks older than 3.9 Ga exist. The most direct evidence from the Hadean era ($\geq$4.0 Ga) comes from submillimeter detrital zircon grains such as those identified in 3.3 Ga metasediments found in the Jack Hills of Western Australia, some of which date back to 4.4 Ga (Harrison 2009). They should be interpreted with caution, however, as the number of grains older than 3.9 Ga is limited, and there is some debate as to whether the zircons crystallized from wet, minimum-melt granites or impact-produced melt sheets (Harrison 2009, Kenny et al. 2016). These zircons have a broad abundance peak between 3.9 and 4.2 Ga, with a maximum at 4.1 Ga (Blichert-Toft & Albarède 2008, Holden et al. 2009). Marchi et al. (2014) argued that the age distribution of the Jack Hills zircons could correspond to an upick in the bombardment flux through crustal thickening associated with impact ejecta and eutectic melting of buried wet crustal material. Possible links between thermal events in these zircons and impacts at $\sim$3.9 Ga have been proposed (Trail et al. 2007, Abbott et al. 2012, Bell & Harrison 2013).

An oft-neglected constraint comes from terrestrial impact spherule beds. When a large impactor strikes the Earth, it produces a vapor-rich ejecta plume containing numerous sand-sized melt droplets, most of which rise above the atmosphere. Eventually the droplets cool and fall back, forming a global layer that can be several millimeters to many centimeters thick for Chicxulub-sized or larger impact events (Simonson & Glass 2004, Bottke et al. 2012, Johnson & Melosh 2012, Lowe et al. 2014). These layers potentially tell us about large ancient impact events, even if the crater formed has been lost to erosion or tectonics. Multiple spherule beds have been found in Archean and early Proterozoic terrains, with the oldest spherule deposits at 3.47 Ga (see Section 3). Modeling their formation, Bottke et al. (2012) found that their age distribution likely corresponds to 70–80 craters with $D > 150$ km forming on Earth between 1.7 and 3.7 Ga. Given that the likely impact ratio between the Earth and Moon is $\sim$20, this number would explain the four such craters...
on the Moon formed over 3–3.7 Ga. Collectively, they suggest the LHB had a long-lived tail that potentially lasted down to ~2 Ga on Earth for Chicxulub-sized impact events (see Section 2.3).

Comparable LHB tails should also exist on Mercury, the Moon, and Mars, although because they are smaller targets, they tend to sample smaller LHB projectiles. Thus, there is no single termination time for the LHB, such as the ~3.8 Ga time often given in textbooks. Instead, the LHB likely ends at different times for different projectile sizes on different worlds.

3. PLAUSIBLE DYNAMICAL MODELS FOR EARLY BOMBARDMENT

Here we review several dynamical scenarios capable of producing early bombardment among the terrestrial planets (see Morbidelli et al. 2015 for additional details). For considerations of space, we only concentrate on a few possibilities, although some scenarios not discussed may have intriguing elements (e.g., bombardment from the disruption of a large asteroid, debris produced when large protoplanets strike the Earth or Mars, the destabilization of a putative terrestrial planet that may affect the asteroid belt; Chambers 2007, Čuk 2012, Bottke et al. 2015, Minton et al. 2015a).

Four main sources of potential impactors can be linked to planet formation processes. The first is a population of planetesimals left over from accretion (Morbidelli et al. 2001, Bottke et al. 2007). Dynamically excited by protoplanets and planetary embryos, these bodies may survive long after the planets have reached their full sizes. The second is a population of objects that escape from the young main asteroid belt (Walsh et al. 2011, Morbidelli et al. 2015, Nesvorný et al. 2017). Indeed, we expect the flux of material out of the main belt to be larger than it is today, particularly if the belt was dynamically excited. This is because many of the most dynamically unstable regions of the belt would have been populated initially, whereas today they are empty. The third is a hypothetical population of objects that formed near/in the asteroid belt or beyond Saturn’s orbit that was left on very high eccentricity, Earth-crossing orbits by early giant planet migration while the solar nebula still existed (Walsh et al. 2011). This population could be considered an add-on to leftover planetesimals. These scenarios would all produce monotonically decreasing impactor populations (declining bombardments). The length of their bombardment tails would depend on the planet formation model invoked, but some should be able to produce large impacts on the terrestrial planets for hundreds of millions of years or more.

The fourth is related to migration of the giant planets after the solar nebula had dissipated. During the last decade, it has become increasingly apparent that the giant planets did not form on their current orbits. In particular, the structure of the Kuiper belt beyond Neptune shows strong evidence that Neptune was once much closer to the Sun (Malhotra 1993, Thommes et al. 1999, Nesvorný & Vokrouhlický 2016). The likely mechanism for this migration involves the gravitational depletion of a massive population of small bodies by the giant planets (Fernández & Ip 1984, Thommes et al. 1999, Tsiganis et al. 2005). Such planetary migration will affect the impact rate in the inner Solar System because it would liberate previously stable small body populations. A quantified and well-developed description of this behavior is provided by the Nice model (Gomes et al. 2005, Tsiganis et al. 2005) (Figure 8). The Nice model is an umbrella term for a broad class of dynamical models in which the giant planets experienced a dynamical transition. It is potentially powerful because it explains not only the current orbits of the giant planets but also the dynamical state of small body populations across the Solar System. For space reasons, we refer the interested reader to Morbidelli et al. (2015), Roig & Nesvorný (2015), Nesvorný & Vokrouhlický (2016), Roig et al. (2016), Vokrouhlický et al. (2016), and Nesvorný et al. (2017) for a list of recent Nice model references and applications.

In the Nice model, the giant planets are assumed to have formed in a more compact configuration between 5 and ~17 AU. They were surrounded by a primordial disk of small icy planetesimals
Planetary orbits and disk particle positions in the Nice model, showing four different times from a reference simulation. In this run, the giant planets were initially on nearly circular, coplanar orbits with semimajor axes of 5.5, 8.2, 11.5, and 14.2 AU. The initial disk contained 35 Earth masses between 15.5 and 34 AU. (a) Beginning of planetary migration. (b) Just prior to the dynamical instability of the giant planets. (c) During the instability. (d) 200 Ma later, with the final planet orbits. The timing of the instability is unknown. This event also triggers the loss of asteroids from the primordial asteroid belt.

(i.e., comets), the ancestor of today’s Kuiper belt, residing between ~20 and 30 AU. The net mass of the disk was several tens of Earth masses. Slow migration of the planets was induced by gravitational interactions with these icy planetesimals or collisionally produced dust leaking out of the disk. Eventually, this triggered an instability that led to a violent reorganization of the outer planets. Uranus and Neptune were scattered into the primordial disk, flinging its members throughout the Solar System. The instability also drove resonances across the asteroid belt, driving substantial portions of it onto planet-crossing orbits. The majority of impacts on outer Solar System worlds would have been from comets, whereas those on the terrestrial planets and Moon would have been from both asteroids and those comets that survived passage into the inner Solar System.

Intriguingly, the Nice model instability could occur after a delay that may be as short as several tens of millions of years (early instability) or as long as many hundreds of millions of years (late instability). Thus, for the latter case, it may be capable of producing a late uptick in impacts, as suggested in several data sets discussed previously. The timescale of bombardment is short for comets, most of which are lost from the primordial disk over tens of millions of years, and very long for asteroids, which are excited and depleted from the primordial main belt over hundreds of millions of years or more. The asteroid bombardment tail may explain the timing of late-forming $D > 150$ km craters on the Earth (Bottke et al. 2012; for a contrasting view, see Minton et al. 2015a,b; Johnson et al. 2016) (Figure 9).

At present, modeling work suggests that early instabilities are easier to obtain than late ones, although much depends on the precise nature of the trigger event. In addition, reconfiguration of the outer Solar System may drive giant planet resonances across the orbits of the terrestrial planets, potentially producing orbits different from those observed (Agnor & Lin 2012, Nesvorný & Morbidelli 2012, Kaib & Chambers 2016). A proposed solution would be to trigger the instability before planet formation was complete; this would excite protoplanets and planetesimals in the terrestrial planet region, but perhaps not enough that the terrestrial planets that eventually form violate the orbital constraints of our Solar System. On the other hand, for some sets of initial conditions, resonance sweeping driven by late giant planet migration appears fully capable of reproducing the dynamical properties of the terrestrial planets (Roig et al. 2016). We find it remarkable that any solution is possible. This could suggest that a late instability is not something
Asteroid impactors striking the Earth and Moon from a version of the Nice model (Figure 8) that starts at 4.1 Ga, according to Bottke et al. (2012). The plotted impactors are assumed to make \( D > 150 \) km craters. The gray boxes are time intervals with constraints. For the lunar Late Imbrian era (box 1) and Eratostenian era (box 2), there are three to four and zero to one craters of this size observed, respectively. Boxes 3–5 correspond to terrestrial impact spherule beds from specific Archean and early Proterozoic terrains that have been extensively searched: at least ten beds between 3.23 and 3.47 Ga (box 3), four beds between 2.49 and 2.63 Ga (box 4), and one bed between 1.7 and 2.1 Ga (box 5). Over the same time intervals, the model results are essentially the same: For boxes 1 to 5, the model predicts 3 ± 2, 1 ± 1, 9 ± 3, 3 ± 2 and 1 ± 1, respectively.

to avoid but instead is actually required to reproduce the observed orbits of the terrestrial planets. Further work on this important issue is warranted.

4. SYNTHESIS

The dynamical scenarios in Section 3 have profoundly influenced our ideas about the impact rate at epochs earlier than \( \sim 3 \) Ga, and we must take them into account if we are to probe this period of time. Indeed, it is possible or even likely that all four impactor sources are needed to explain the bombardment record of the terrestrial planets and asteroid belt. Determining which sources dominate at different times, however, strongly depends on the assumptions made and models used. For this reason, here we examine different bombardment scenarios in a conceptual manner by combining the impactor sources with a mix and match application of the available constraints. Issues like the relative size of the impactor populations are ignored for simplicity.

The main constraint used in these conceptual models is the ratio of \( D > 150 \) km craters found on the oldest lunar and martian surfaces. From our analysis in Section 2.6, and assuming Mars and the Moon have crater SFDs with fairly similar shapes (which may not be true), we estimated that the Mars to Moon ratio of such impacts is \( \sim 1, 2.8, \) or \( 5.3 \), although intermediate values are also plausible. Modeling results indicate the ratio of leftover planetesimals hitting Mars and the Moon is \( \sim 2:1 \) (e.g., Bottke et al. 2007, Walsh et al. 2011), whereas the combined asteroid and comet components of the Nice model produce a ratio of \( \sim 7–10:1 \) (Bottke et al. 2012, Nesvorný et al. 2017). We apply these values below.
We must also consider crater or basin retention ages of the oldest lunar and martian surfaces. For example, the Moon’s oldest surface could be as young as 4.35 Ga, which may date a global magmatic event (Borg et al. 2015), or as old as ~4.4–4.5 Ga, the putative age of the Moon itself (Touboul et al. 2009, Bottke et al. 2015). For Mars, it is possible that the oldest surfaces correspond to the age of the Borealis basin, which defines Mars’s global topography (Andrews-Hanna et al. 2008). The available evidence suggests this basin formed >4.5 Ga (Nimmo et al. 2008, Borg et al. 2016, Bottke & Andrews-Hanna 2017; see Section 2.5), but a younger formation age and more recent resurfacing cannot be ruled out. It is also possible that large basins would not be retained for some interval after Borealis formation, although this time period may be as short as a few tens of millions of years (Bottke & Andrews-Hanna 2017).

4.1. Scenario 1: Dominantly Declining Bombardment

If we assume that leftover planetesimals dominate the impactor populations striking early Mars and the Moon, and that the crater retention age of both worlds was virtually the same (e.g., both 4.5 Ga), the following pros and cons can be developed.

One pro is that the impact ratio is nearly 2, a value broadly consistent with a Mars/Moon ratio of either ~1 or ~2.8. This would argue that there was no substantial source of late impactors. Accordingly, the giant planets should have experienced an early Nice model–like instability, such that most putative impactors (derived from the destabilized portions of the primordial main belt and outer Solar System disk of planetesimals) were eliminated prior to the crater retention ages of these worlds and/or that their contribution to the net impactor population striking Mars and the Moon was limited. Another appealing aspect to this conceptual model is its simplicity. For example, a 2:1 bombardment ratio readily explains why two $D > 2,000$ km basins exist on Mars (Utopia and Hellas) while one exists on the Moon (South-Pole Aitken basin).

One con of this scenario is that models of leftover planetesimal bombardment often produce many more early impactors than are observed (Bottke et al. 2007). To avoid this issue, crater retention ages for Mars and the Moon would need to be relatively young and similar to one another (e.g., both near 4.3 Ga). Another disadvantage is that leftover planetesimal bombardment models follow a declining bombardment, which would be inconsistent with a paucity of impacts between ~4.1 and 4.4 Ga inferred from meteorite shock ages and lunar impact melts (unless the paucity of lunar melts is a bias produced by a young crater retention age). A third problem is that it is currently unclear whether leftover planetesimals can produce the $D > 150$ km craters that formed on the Earth and Moon between 2 and 3.7 Ga (Sections 2.4 and 2.5) (Bottke et al. 2007, 2012). Finally, this model cannot reproduce a Mars/Moon ratio of 5.3; if that is the true value, some other bombardment component is needed.

4.2. Scenario 2: Dominantly Instability-Driven Bombardment

Here we assume all lunar and martian basins come from a Nice model–like bombardment. Comets would hit first over a few tens of millions of years, and asteroids would produce the rest over a much longer interval. To explain all large impacts, the instability would need to take place prior to the crater retention ages of the lunar and martian surfaces. Depending on the timing of the instability, this could look like a declining bombardment or something akin to a Terminal Cataclysm (Figure 3). An advantage of this model is that mass loss from a dynamically perturbed asteroid belt is potentially the best source of a long bombardment tail (Bottke et al. 2007, 2012). This might allow it to produce $D > 150$ km craters on the Earth and Moon between 2 and 3.7 Ga (Figure 9) (Sections 2.4 and 2.5). However, one disadvantage of this scenario is that none of the Mars/Moon impact ratios (1, 2.8, or 5.3) will work if Mars and the Moon have equal-age surfaces. The only
way out would be to invoke a younger crater retention age for Mars than the Moon, perhaps via a relatively young Borealis formation event or later resurfacing. Moreover, as in Scenario 1, this model is inconsistent with a paucity of impacts between 4.1 and 4.4 Ga inferred from lunar melt and meteorite shock ages.

4.3. Scenario 3: Hybrid Model

Here it is assumed that there are early and late components of bombardment. The early one would be a relatively short-lived declining bombardment from leftovers initiated during the planet formation era, with most impactors gone by \( \sim 4.4 \) Ga. The later one would come from a Nice model–like late instability, although solutions for early instabilities are also intriguing. The late instability would need to start at \( \sim 4.0 \) Ga to perhaps 4.2 Ga in order to be consistent with the distribution of lunar melt rocks and meteorite shock degassing ages.

One advantage of this scenario is that solutions can be found for Mars/Moon impact ratios of 2.8 and 5.3. For the former value, we find a late instability-driven bombardment would make up 25–40% and 10–15% of Mars’s and the Moon’s bombardment, respectively. For the latter value, it would make up 80–90% and 40–70% of Mars’s and the Moon’s bombardment, respectively. Another advantage is that the hybrid model can potentially match all the constraints discussed above while allowing crater retention ages on Mars and the Moon to be near 4.4–4.5 Ga (see Bottke et al. 2012, Morbidelli et al. 2012).

A disadvantage of the hybrid model is that late instabilities tend to be harder to generate than earlier ones, and they require a solution consistent with the current orbits of the terrestrial planets. If such a solution cannot be found, invoking an early instability might work, although at the expense of violating some constraints in Section 2 (e.g., meteorite shock ages). Moreover, the model runs into problems if the Moon’s crater retention age is substantially younger than Mars. Finally, this model cannot explain a Mars/Moon impact ratio of \( \sim 1 \) without strongly modifying the crater retention ages of the Moon and Mars.

5. CONCLUSIONS

A common assertion, often overheard in the hallways of planetary science meetings, is the phrase “I do not believe in the Late Heavy Bombardment.” It is not clear precisely what this means, but it likely refers to doubts that the Moon and other worlds were hit by a spike of large impact events between \( \sim 3.7 \) and \( \sim 3.9 \) Ga. Given the evidence provided here, we agree that the original basis for a strong version of the Terminal Cataclysm hypothesis has been substantially weakened.

With this said, however, it is worth considering that two nearly 1,000 km lunar basins, Imbrium and Orientale, formed on the Moon during this short interval. Given the gravitational cross sections of the Earth and Moon, this implies that nearly 40 such colossal events took place on Earth, possibly over a comparable timescale. There is also compelling evidence that heavy bombardment continued on Earth and the Moon well after this time, perhaps on Earth all the way to \( \sim 2 \) Ga. When size distributions are also considered, it is unavoidable that at least some Archean-era impacts on Earth may have been comparable to Orientale-formation events on the Moon. Now that is a late heavy bombardment!

At present, the evidence is at least modestly supportive of a hybrid model where two bombardment phases take place, an early one that ends by \( \sim 4.4 \) Ga produced by leftover planetesimals and a second one initiated by late giant planet migration that starts near \( \sim 4–4.2 \) Ga. This scenario will produce a long bombardment tail, and it is fairly viable within existing dynamical modeling work. Nonetheless, late bombardment episodes driven by delayed giant planet dynamical instabilities are
more difficult to get to work, and specific conditions may be needed to avoid modifying the orbits of the terrestrial planets (although this might turn out to be a highly desirable trait; see Section 3). The ratio of large craters on Mars versus the Moon may also favor a declining bombardment, although much depends on how this record is debiased on both worlds and whether their crater SFDs indeed have similar shapes.

**FUTURE ISSUES**

We suggest progress in the following areas would help us to better understand the nature of the LHB.

1. Absolute ages of the most ancient lunar and martian surfaces and basins would provide direct constraints on the timing of the LHB. New sample return missions will be needed to obtain those data.

2. Minerals and rocks from terrestrial Hadean and Archean terranes provide cost-effective sampling of the early Earth that can produce new constraints on early bombardment.

3. Nature provides us with an abundance of extraterrestrial samples from asteroids, the Moon, and Mars via meteorites. Many of these can tell us about ancient impact events or planetary evolution. There is also much to be learned from existing samples using new ideas and instrumentation.

4. New observations and modeling of small body populations and impact processes would inform studies of small bodies in both the asteroid belt and the trans-Neptunian populations, both of which contributed to the LHB, as well as the projectile sizes needed to make large craters under specific conditions.

5. Considerably more physics could be added to existing numerical models of planet formation and migration, disk evolution, and small body evolution, e.g., planetesimal and planet formation models that accurately include gas disk processes, collisional evolution among planetesimal populations, and planetesimal self-gravity included in massive disks. Ultimately, the LHB is linked to planet formation and migration; both need to be solved together.

**DISCLOSURE STATEMENT**

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

**ACKNOWLEDGMENTS**

We thank J. Andrews-Hanna, R. Canup, C. Chapman, B. Cohen, M. Kirchoff, D. Kring, H. Levison, S. Marchi, A. Morbidelli, D. Nesvorný, A. Parker, S. Robbiins, J. Salmon, T. Swindle, and D. Vokrouhlický for numerous useful and stimulating discussions. We also credit SSERVI’s “Early Bombardment of the Solar System” workshops for illuminating so many of the key issues. W.F.B.’s participation was supported by NASA’s SSERVI program Institute for the Science of Exploration Targets (ISET) through institute grant NNA14AB03A. M.D.N. would like to thank Vickie Bennett for comments on the manuscript, and G.J. Wasserburg, who had the answers in 1967.
LITERATURE CITED


Boehnke P, Harrison TM. 2016. Illusory late heavy bombardments. PNAS 113(39):10802–6


Brinicombe P, Harrison TM. 2016. Illusory late heavy bombardments. PNAS 113(39):10802–6


Minton DA, Jackson AP, Asphaug E, Fassett CI, Richardson JE. 2015a. Debris from Borralis basin formation as the primary impactor population of Late Heavy Bombardment. Presented at Early Sol. Syst. Impact Bombard. III, Feb. 4–5, Houston, TX

Minton DA, Richardson JE, Fassett CI. 2015b. Re-examining the main asteroid belt as the primary source of ancient lunar craters. Icarus 247:172–90

Morbidelli A, Bottke WF, Nesvorný D, Levison HF. 2009. Asteroids were born big. Icarus 204(2):558–73


Contents

Researching the Earth—and a Few of Its Neighbors
Susan Werner Kieffer ................................................................. 1

The Fascinating and Complex Dynamics of Geyser Eruptions
Shaul Hurwitz and Michael Manga ............................................. 31

Plant Evolution and Climate Over Geological Timescales
C. Kevin Boyce and Jung-Eun Lee ............................................... 61

Origin and Evolution of Water in the Moon’s Interior
Erik H. Hauri, Alberto E. Saal, Miki Nakajima, Mahesh Anand,
Malcolm J. Rutherford, James A. Van Orman, and Marion Le Voyer ........ 89

Major Questions in the Study of Primate Origins
Mary T. Silcox and Sergi López-Torres ..................................... 113

Seismic and Electrical Signatures of the Lithosphere–Asthenosphere
System of the Normal Oceanic Mantle
Hitoshi Kawakatsu and Hisashi Utada ......................................... 139

Earth’s Continental Lithosphere Through Time
Chris J. Hawkesworth, Peter A. Cawood, Bruno Dhuime, and Tony I.S. Kemp .... 169

Aerosol Effects on Climate via Mixed-Phase and Ice Clouds
T. Storelvmo .............................................................................. 199

Hydrogeomorphic Ecosystem Responses to Natural and
Anthropogenic Changes in the Loess Plateau of China
Bojie Fu, Shuai Wang, Yu Liu, Jianbo Liu, Wei Liang, and Chiyuan Miao ........ 223

Interface Kinetics, Grain-Scale Deformation, and Polymorphism
S.J.S. Morris ............................................................................ 245

Back-Projection Imaging of Earthquakes
Eric Kiser and Miaki Ishii ............................................................. 271

Photochemistry of Sulfur Dioxide and the Origin of Mass-Independent
Isotope Fractionation in Earth’s Atmosphere
Shubei Ono .............................................................................. 301

Southeast Asia: New Views of the Geology of the Malay Archipelago
Robert Hall ............................................................................ 331
Forming Planets via Pebble Accretion
Anders Johansen and Michiel Lambrechts .................................................. 359

Tungsten Isotopes in Planets
Thorsten Kleine and Richard J. Walker ...................................................... 389

Shape, Internal Structure, Zonal Winds, and Gravitational Field of Rapidly Rotating Jupiter-Like Planets
Keke Zhang, Dali Kong, and Gerald Schubert ............................................. 419

Effects of Partial Melting on Seismic Velocity and Attenuation: A New Insight from Experiments
Yasuko Takei .............................................................................................. 447

Origin and Evolution of Regional Biotas: A Deep-Time Perspective
Mark E. Patzkowsky .................................................................................... 471

Statistics of Earthquake Activity: Models and Methods for Earthquake Predictability Studies
Yosibko Ogata ............................................................................................ 497

Tectonic Evolution of the Central Andean Plateau and Implications for the Growth of Plateaus

Climate and the Pace of Erosional Landscape Evolution
J. Taylor Perron ........................................................................................... 561

The Rise of Animals in a Changing Environment: Global Ecological Innovation in the Late Ediacaran
Mary L. Droser, Lidya G. Tarhan, and James G. Gebling ................................ 593

The Late Heavy Bombardment
William F. Bottke and Marc D. Norman .................................................... 619

Reconstructing Climate from Glaciers
Andrew N. Mackintosh, Brian M. Anderson, and Raymond T. Pierrebumbert ...... 649

Autogenic Sedimentation in Clastic Stratigraphy
Elizabeth A. Hajek and Kyle M. Straub ...................................................... 681

Errata
An online log of corrections to Annual Review of Earth and Planetary Sciences articles may be found at http://www.annualreviews.org/errata/earth

x  Contents